Marcin Michalski

WRITTEN MOROCCAN ARABIC

A Study of Qualitative Variational Heterography

Wydawnictwo
Naukowe UAM

Written Moroccan Arabic

UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
SERIA ORIENTALISTYKA NR 7

Marcin Michalski

Written Moroccan Arabic

A study of qualitative variational heterography

WYDAWNICTWO NAUKOWE

POZNAŃ 2019

Recenzent: prof. Ignacio Ferrando

Publikacja sfinansowana przez
Rektora Uniwersytetu im. Adama Mickiewicza w Poznaniu, Wydział Neofilologii UAM
oraz Instytut Językoznawstwa

Wydano na podstawie maszynopisu gwarantowanego

Autorzy fotografii na okładce: Joanna Nowak-Michalska i Marcin Michalski

Projekt okładki: K. & S. Szurpit
Korekta językowa: Christopher Whyatt
Redakcja techniczna: Dorota Borowiak
Łamanie komputerowe: Marcin Tyma

ISBN 978-83-232-3396-1
ISSN 1730-8771

WYDAWNICTWO NAUKOWE UNIWERSYTETU IM. ADAMA MICKIEWICZA W POZNANIU
61-701 POZNAŃ, UL. FREDRY 10
www.press.amu.edu.pl
Sekretariat: tel. 61 829 46 46, faks 61 829 46 47, e-mail: wydnauk@amu.edu.pl
Dział Promocji i Sprzedaży: tel. 61 829 46 40, e-mail: press@amu.edu.pl

Wydanie I. Ark. wyd. 13,00. Ark. druk. 13,50

DRUK I OPRAWA: VOLUMINA.PL DANIEL KRZANOWSKI, SZCZECIN, UL. KS. WITOLDA 7-9

Contents

Abbreviations and symbols

Av	asymmetrical variation
CA	Classical Arabic
def	definite
du	dual
EA	Egyptian Arabic
f	feminine
Fv	free variation
lit.	literal translation
MA	Moroccan Arabic
obl	oblique case
pl	plural
Pv	Pseudo-variation
Rv	Restricted variation
SA	(Modern) Standard Arabic
‹›	contain transliteration
()	contain irrelevant distinctive graphic elements
[]	contain accidentals (graphic elements responsible for irrelevant semantic differences)
{}	contain expected variants and expected variations (not recorded in the corpus)
*	marks non-existing (not recorded and not expected) forms
<	in transcription, indicates the underlying morphological structure

Arabic words that have no gloss (morphological description) are masculine singular and – in the case of nouns, adjectives and participles – indefinite.

A key to the abbreviations of the titles of the literary works cited as sources is given in References, under Primary sources.

Transliteration and transcription

Transliteration consists in converting the graphs of Arabic script into Latin letters, if needed with diacritics, or, in rare cases, into special signs, with a one-to-one correspondence in both directions (biuniqueness). Transliteration, whether of separate graphs or graphic words, is given in angle brackets: ‹›. It is used in this book alongside the Arabic notation to ensure the precision of description on the one hand and to facilitate the reading to those that do not have a good command of Arabic, on the other hand.

The transliteration system proposed for the present book uses mostly symbols borrowed from the Romanization systems used in Arabic Studies for SA, modified and expanded with the addition of some other signs[1]. The transliteration symbols are generally intended to reflect the most typical phonetic value of a given graph in SA or MA or, in less frequent cases, its shape. 42 primary graphs are distinguished. They include the 28 elements of the Arabic alphabet, for which the term 'letter' is reserved, the bare *hamza* sign: ء ‹o›, and modifications of these letters consisting in adding the *hamza* or the *madda* signs and changing the number and/or position of the dots and strokes. The graphs are given in Table 1, with their four forms depending on their graphic environment, i.e. whether they are or are not connected to the preceding and following graph[2].

[1] A complex system for transliterating Arabic script with special characters is proposed and used in Mumin & Versteegh (2014: 12-21). However, the symbols used there to reflect the writing practices in numerous African languages would lack transparency in a study devoted to MA.

[2] Since some graphs (non-connectors) are never connected to the following graph, even in the word-medial position, and because the graph ‹o› is never connected on either side, the factor determining the shape of a graph is the *environment*, and not its *position in a word*, as is misleadingly stated in some presentations of Arabic script. Hence, the terms 'initial', 'medial' and 'final' used here refer to the position in a group of letters connected with one another. Coulmas (2003: 123) calls it a 'writing group' and remarks that it does not always concide with a graphic word. For instance, the graph ل ‹l› is *word-medial* in the SA graphic word ملء ‹mlo› *malʔun* '(the act of) filling', but is *final* in its writing group.

Table 1. *The transliteration system*

Isolated form	Non-isolated forms			Transliteration
	Final	Medial	Initial	
ا	ـا	ـا	ا	‹a›
أ	ـأ	ـأ	أ	‹å›
آ	ـآ	ـآ	آ	‹ã›
إ	ـإ	ـإ	إ	‹ą›
ب	ـب	ـبـ	بـ	‹b›
پ	ـپ	ـپـ	پـ	‹p›
ت	ـت	ـتـ	تـ	‹t›
ٹ	ـٹ	ـٹـ	ٹـ	‹ṭ›
ج	ـج	ـجـ	جـ	‹ǧ›
ح	ـح	ـحـ	حـ	‹ḥ›
خ	ـخ	ـخـ	خـ	‹x›
چ	ـچ	ـچـ	چـ	‹č›
د	ـد	ـد	د	‹d›
ذ	ـذ	ـذ	ذ	‹ḏ›
ر	ـر	ـر	ر	‹r›
ز	ـز	ـز	ز	‹z›
س	ـس	ـسـ	سـ	‹s›
ش	ـش	ـشـ	شـ	‹š›
ص	ـص	ـصـ	صـ	‹ṣ›
ض	ـض	ـضـ	ضـ	‹ḍ›
ط	ـط	ـطـ	طـ	‹ṭ›
ظ	ـظ	ـظـ	ظـ	‹ẓ›
ع	ـع	ـعـ	عـ	‹ʕ›
غ	ـغ	ـغـ	غـ	‹ġ›
ف	ـف	ـفـ	فـ	‹f̣›
* ف (ف or ق)	ـف (ـف or ـق)	فـ	فـ	‹f› (‹f$_{ị}$› or ‹f$_{q}$›)
ق	ـق	ـقـ	قـ	‹q›
* ڤ (ڤ or ڤ)	ـڤ (ـڤ or ـڤ)	ڤـ	ڤـ	‹v› (‹v$_{ị}$› or ‹v$_{q}$›)
ك	ـك	ـكـ	كـ	‹k›
گ	ـگ	ـگـ	گـ	‹ǵ›
* ڴ	*ـڴ	ـڴـ	ڴـ	‹ĝ›

ل	ـل	ـلـ	لـ	‹l›
م	ـم	ـمـ	مـ	‹m›
ن	ـن	ـنـ	نـ	‹n›
ه	ـه	ـهـ	هـ	‹h›
ة	ـة	/	/	‹ħ›
و	ـو	ـو	و	‹w›
ۇ	ـۇ	ـۇ	/	‹ẘ›
ى	ـى	/	/	‹á›
ي	ـي	ـيـ	يـ	‹y›
ئ	ـئ	ـئـ	/	‹ẙ›
ء	ء	ء	/	‹o›

Notes

A slash / means that a graph is not used in a given position.

* For the discussion of these graphs, see Sect. III.1.

Since the graph پ, i.e. a ش ‹š› with three extra dots below, the use of which has been recorded in MA (Lerchundi 1900: 4 [1872: 5, footnote 1]), does not occur in the texts examined within the framework of the present study nor has been described in recent scholarly literature, it is not included in the table.

The vocalization signs, i.e. the secondary graphs, are generally considered irrelevant for the purposes of this study. They are transliterated as superscript vowels reflecting their function in SA: The *fatḥa*, ـَ, marking short *a* in SA, is transliterated as ‹^{a}›, the *kasra*, ـِ , marking short *i*, is transliterated as ‹^{i}› and the *ḍamma*, ـُ, marking short *u*, is transliterated as ‹^{u}›. The *šadda*, which in SA marks the lengthening (gemination) of a consonsant or semivowel, is transliterated as ‹ˑ› following a given transliteration symbol. The *sukūn*, which in SA marks the absence of a vowel after a consonant, but in MA can also mark *ə* (cf. Aguadé 2006: 256), is tranliterated as a superscript bullet ‹˙› following a given transliteration symbol. The *tanwīn*, i.e. ـً , ـٍ or ـٌ, is transliterated as a superscript ‹^{an}›, ‹^{in}› or ‹^{un}›, respectively. Space is transliterated as a low line: ‹_›. For instance, دايماً *daymən* ‘always’ is transliterated as ‹dayma^{an}›; هاذ الْغُرَّاف *had l-ġŭṛṛaf* ‘this water jar’ is transliterated as ‹ha_d_al˙ġ^{u}rˑ^{a}af›.

Apart from transliteration, broad transcription, combining phonological and morphological representations, is given for every SA and MA graphic word, in italics, with hyphens separating affixed units (such as the definite articles, suffixed pronouns, some prepositions, particles and conjunctions). For MA, the symbols and notation follow those used in Aguadé & Benyahia (2005). Some of the symbols differ significantly from their corresponding IPA characters: *t̲* = [θ],

ǧ = [ʤ], ž = [ʒ], č = [ʧ], ḥ = [ħ], ḏ = [ð], š = [ʃ], ġ = [ɣ], y = [j]. A dot bellow a letter, except for ḥ, signals pharyngealization, e.g. ṣ = [sˤ], ẓ = [ʒˤ], etc.[3]. The symbols ā, ī and ū, used in transliterating SA, represent long vowels. Reflecting some kind of an idealized spoken norm, the transcription system adopted in the present study is thus a systematic transcription, which arguably is sufficient for our present purposes. An alternative choice, i.e. using impressionistic transcription, which would reflect all of the phonetic features of a given phonetic word, has been rejected for a number of reasons. First, a graphic word may represent more than one phonetic word. Second, certain phonetic features, such as affricatization or the exact phonetic value of some vowels, are never reflected in writing and are irrelevant for our considerations. Finally, in most cases, phonetic transcription is not necessary for elucidating the aspects under analysis. It is supplied, in square brackets, [], only in special cases, if the phonetic form of an element under discussion is at odds with the broad transcription used[4].

In quoting Arabic data from other authors, the original notation is retained, except for those taken from Harrell (1962): four of the transliteration symbols used there are replaced as follows: *e* > *ə*, *o* > *ŭ*, ع > ʕ and ح > *ḥ*.

As a rule, graphic words are analyzed and represented regardless of their graphic environments. However, if there is a possibility that the graphy of a given graphic word is influenced by a neighbouring one, this is taken into consideration.

The full description of every graphic word used as an example comprises: its (photo)graphic reproduction, transliteration, abbreviation – in parentheses – indicating its source (see References, under Primary sources) with the page and line number (the superscript means 'the *n*-th line from the top', the subscript – 'the *n*-th line from the bottom'), transcription and English translation. Whenever necessary, the heterograph represented on the left is referred to as Form *A*, or simply *A*, and the one represented on the right as Form *B*, or simply *B*.

---

[3] Pharyngealization is one of the two phonetic manifestations of emphasis in MA, the other one being velarization (cf. e.g. Caubet 2008: 275).

[4] For a detailed presentation of the sounds of MA, see Aguadé (2003).

Introduction

For some time now, writing activity in MA[5] has been experiencing relatively considerable progress. Moroccans communicate more and more in written form in non-formal situations, via the Internet and mobile phones, and they seldom do this in SA. Written MA is also used for advertising purposes and in the press, both printed and electronic. Since the middle of the 20th century, an increasing amount of printed literary texts has been being published in MA, although this growth is by far not as intensive as in Egypt, where books written in the local variety of Arabic are easily available nearly in any modern bookshop. It is this type of text, more specifically prose and theatre pieces, that will be of interest to us here.

Putting aside the issue of MA being written with the use of the Latin alphabet[6], the present study is devoted to how it is written with the use of Arabic script, used primarily for writing SA. In this respect, SA, the official, formal and prestigious variety of Arabic, has the status of a donor language. Since MA, not unlike other Arabic dialects[7], is considered as low prestige and thus unfit for being a vehicle of formal written communication, its spelling has not been standardized by any official body. Instead, it is created, often collectively, by authors, editors and publishers, who, in this specific function, will be jointly referred to as 'spellers'[8].

Since the phonetics, phonology, morphology and vocabulary of MA differ, sometimes significantly, from those of SA, the orthographic principles of the latter cannot always be easily adopted for writing in the former. Therefore, they are often modified, disregarded or deliberately violated. The lack of generally acknowledged spelling rules for MA – in contrast to SA, the orthography of which is governed by relatively strict norms – and the fact that authors writing in MA create their own spelling rules but often do not feel obliged to stick to them throughout the entire text makes its graphy unstable and fraught with variational heterography

---

[5] A term often used in both scholarly and popular literature to refer to MA is *d-dariža* (*darija* in English and French), a word borrowed from SA الدارجة (اللغة) *(al-luġa) ad-dāriğa* 'the colloquial (language)'.

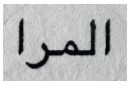
[6] Like many other important Arabic dialects, MA is often written in Latin script in the Internet and SMS communication (Caubet 2004, 2012, 2013; Moscoso 2009), but sometimes also in the press (Benítez Fernández 2012a). The translation *Hhi*. is a rather isolated example of a literary work written in MA with the use of this script.

[7] Maltese, with its special history, structure and status of a national language, is a notable exception.

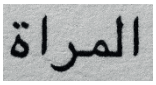
[8] I am using this term in this meaning after Carney (1994).

– a phenomenon consisting in more than one graphic word corresponding to one phonetic word (see Sect. III.2.1. for details). Variational heterography, a pervasive trait of MA spelling, can occur even in a text written by a single author, on a single page and in a single line. Example (1) shows a pair of heterographs occurring in one line.

(1)



‹almra› (*N* 30_{8})
lə-mṛa ‘woman^{def}’

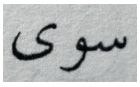


‹almra**ħ**› (*N* 30_{8})

In (2), two different graphies are used for the same phonetic word repeated in a correlative conjunction.

(2)

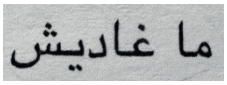
‹sw**a**›
swa… swa… ‘either… or…’



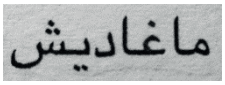
‹swá› (*DB* $212_{2\text{-}3}$)

Graphic multiplets, i.e. sets of more than two different graphic words for one phonetic word, are easy to find in MA, even within a single text. For instance, the negated future auxiliary *ma ġadi-š* has three heterographs in three adjacent lines in *N*:

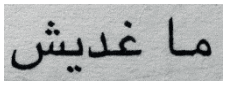
(3)



‹ma_ġadyš› (*N* 29^{12})



‹maġadyš› (*N* 29^{13})

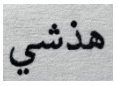


‹ma_ġdyš› (*N* 29^{14})

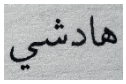
ma ġadi-š ‘not going to’ (negated future marker)

A yet greater number of heterographs can be identified if various sources are compared. Below, eleven heterographs of the MA phrase *had š-ši* ‘this’ illustrate how numerous they can be, especially for frequently used words (the vocalization signs are disregarded).

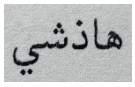
(4)



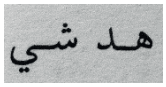
‹ḥd̲šy›
(*QQ* 46_{4})



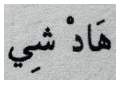
‹hadšy›
(*Ṣa.* 221^{5})



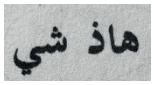
‹haḏšy›
(*ʕD* 107_{8})



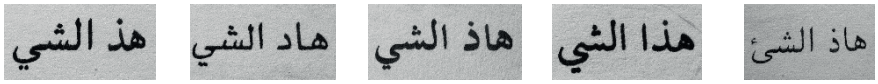
‹hd_šy›
(*T̲M* 53_{3})



‹had_šy›
(*Ḍa.* 55_{4})



‹haḏ_šy›
(*ḤB* 46^{12})



‹hd̲_alšy› ‹had_alšy› ‹had̲_alšy› ‹hd̲a_alšy› ‹had̲_alšẙ›
(*XM* 55[4]) (*Bə.* 43[8]) (*Ḥ B* 47[12]) (*Ḥ D* 21$_1$) (*Ḍ K* 22$_3$)
had š-ši ‘this’, lit. ‘this thing’

Heterographs may differ from one another in more than one respect. The three basic types of heterography are: (i) qualitative, when different graphs are used, (ii) quantitative, when a graph is present in one graphic word and absent in another, and (iii) linear, when one heterograph is written without a space and another contains a space. Example (4) shows that heterography can also be mixed, i.e. combine two or more basic types. And it is very often the case.

The MA graphy is a triply complex macrosystem: First, there is not one Moroccan Arabic dialect but a bundle of Moroccan dialects. Second, there exist a number of different individual spelling microsystems. Third, these are usually internally unstable. Nevertheless, the graphic macrosystem of MA does work: texts written in it are readable without greater difficulties. This suggests that the existence of the area of variation, used by spellers quite extensively and imaginatively, does not result in serious comprehension problems and that apart from these shifting sands there must be a core of stability.

The task of describing such an unstable object may be difficult but it is worth undertaking since we are currently witnessing an intensified crystallization of what in the future may evolve into a system of generally acknowledged spelling norms. In other words, the current graphic unstability of MA is most probably a stage preceding, by an unspecified period of time, standardization. This is a stage which most languages with standardized orthography have passed through. However, even without any standardizing aspirations in mind, the phenomenon of written MA deserves investigation for theoretical linguistic reasons as it represents a natural, spontaneous stage of the development of a writing system in which no regulatory and restrictive powers intervene. Examining it gives us an insight into the relationships between the units of its graphic macrosystem as well as into the linguistic awareness of its users.

The extant studies dealing with how modern Arabic dialects are written are not very numerous and are primarily concerned with how particular sounds are represented in writing (see Sect. II.1.). This issue is also discussed in the introductions to those text books and dictionaries of Arabic dialects in which Arabic script is used. However, these presentations are merely of a practical character and do not tell us much about the structure of the graphic system(s). The aim of the present linguistic study is to provide a systematic non-normative and non-evaluative

description of the graphic (macro)system of MA (used in literary texts), as far as qualitative variation is concerned, and thus to fill, at least partly, a gap in our knowledge about this fragment of the reality of MA, and, by extension, of the Arabic language. My initial intention of providing a complete description of the three types of heterography in MA – qualitative, quantitative and linear – had to be narrowed down to the first type when the scope of research turned out to be too large for one volume. The decision to limit the study to qualitative heterography may seem to restrict the heterography landscape in the horizontal dimension. However, in the vertical one, it allows us a more profound penetration of the problem. The two remaining types of heterography will be hopefully studied and described in the future.

Since the linguistic reality to be described, the MA graphic system, is an abstract entity, it is accessible for empirical examination only indirectly *via* written texts, which are its most tangible manifestation. The description is therefore based on a corpus. It includes literary texts written by native speakers of MA and published within a period of twenty years (1991-2012)[9]. Its character is synchronic, as changes which might have occurred within the MA graphic system during this time are not taken into consideration.

In order to describe this abstract system, the corpus is subjected to an analysis consisting in reading the texts, extracting relevant data, i.e. pairs of heterographs at the level of graphic words, establishing pairs of variants at the level of graphs and finally categorizing them in a systematic way. The theoretical framework proposed for this purpose will be based, to some extent, on structuralist concepts used in phonetics and phonology as well as graphetics and graphemics. Since within graphetics and graphemics, attention has been generally devoted to graphic systems which use Latin script (see, for instance, Günther 1988: 64-98 for German), some existing concepts need to be modified in order to make them capable of capturing relevant phenomena characterizing Arabic script.

The study revolves around two axes. The first axis is the hypothesis that some elements in MA graphic words are characterized by variation while others are not. If this hypothesis is verified, the following questions arise: (1) Which elements show variation and which do not? (2) What are the reasons for variation, i.e. what are the spelling principles underlying the graphy of particular graphs? (3) Which words are graphic invariants and why?[10] (4) How ambiguous is MA graphy? (5)

[9] The year of publication of a text does not necessarily coincide with the year of its composition. This concerns, for instance, theatre pieces, some of which were written and staged in the 1970s (cf. Langone 2008: 57).

[10] A similar question was asked by Hoogland (2013: 68): "What do these words [i.e. words that have only one spelling variant throughout his corpus – M.M.] have in common that would explain this uniformity in their spelling?". However, his answer (according to which these words contain

What system is constituted by relations between the units of MA graphy (graphs)? At present, only qualitative variational heterography and qualitative invariants are investigated, but absolute graphic invariants are obviously the most interesting issue from the theoretical and descriptive point of view. They are also of greatest importance for a possible future standardization of the MA graphy. The second axis is an attempt at establishing an inventory of MA graphemes on the basis of the variational properties of graphs. In order to do this, it needs to be decided what relations between the graphs of MA and what other factors are relevant for classifying particular graphs into graphemes.

The description of such a complex entity as the MA spelling system requires some idealizations and simplifications because not every aspect of the problem can be taken into consideration. One such aspect that could not be properly investigated in this study is the exact frequency of particular graphies, an issue that has not been considered to be of great relevance for our principal concerns. However, whether a form is frequently used or marginal and exceptional is always signalled in the text.

The results of the study are intended to contribute to Arabic linguistics, but also to enrich the general linguistic knowledge about the systems of writing used for various languages. Apart from providing theoretical insights, they may be practically useful for anyone interested in developing and using a systematic graphy of MA (authors writing literature, dictionaries or course books, teachers, etc.).

---

only one long vowel, which might explain "the agreement on the representation of the vowels", and that these words do not have SA cognates, which could inspire disagreement) accounts for only a fragment of the MA graphic reality.

The corpus

The corpus consists of nearly 1900 printed pages representing 32 literary texts written by native users of MA and published in Morocco (with two exceptions which appeared in Lebanon and Syria)[11]. Analyzing a larger number of texts representing one single genre has been preferred to examining a corpus composed of various genres, for instance: press and electronic sources such as the Internet social networking services, chat rooms, Internet forums, etc., in order to avoid the necessity of accounting for possible graphic pecularities of particular text genres. The literary works included in the corpus are both prose and drama texts published in the form of printed books. A list of their titles with their abbreviations used throughout this book is given in References, under The Corpus. Some of these works are written entirely in MA or contain chapters or longer passages in this variety, while others are written mostly in SA but contain isolated MA words, phrases or sentences. Poetic fragments are generally not taken into consideration because of the risk that the form of a given graphic word is a result of *licentia poetica* or stylization. Words occurring in poetic contexts are used on condition that they can be considered free of these influences. Well-established borrowings from other languages, mostly French and Spanish, are treated as belonging to MA, whereas foreign expressions resulting from code-switching or code-mixing transcribed into Arabic script are not analyzed in this work.

The texts included in the corpus are not stylistically and lexically uniform. Some, e.g. *ḤB* or *MX*, represent a very colloquial style, which would correspond to what Youssi refers to as "arabe marocain quotidien" (1992: 25-26). Some others, e.g. *Ṛḥ.*, are written in a more educated variety, situated closer to SA, used in spoken formal situations, avoiding regional features and known from Moroccan radio and television, which Youssi terms "arabe marocain moderne" (1992: 25).

The analysis of written sources of this kind is fraught with particular practical problems. Since there is no digitized and indexed corpus of MA texts that could be used, paper editions analyzed in a traditional way, i.e. read visually, constituted my only source of data. Some of the titles have been digitized by Google (Google

[11] The corpus does not include collections of proverbs. Spelling syśtems used or proposed by both Weśterners and Arabs in scholarly works, such as Otten & Hoogland (1983), al-Fāsī (1986: 25-28), al-Midlāwī (2001) and Durand (2004), as well as in didactic publications, e.g. Hoogland (1996), Benjelloun (1998), Moscoso García (2006) and *Moroccan Arabic* (2011), are a separate phenomenon and are not taken into consideration.

Books) but this tool turned out to be of limited use because it offers, at best, only a snippet view for them, normally with no page number indication and often very poorly OCRed; there is no possibility to purchase a complete digital copy, either. Google Books was an auxiliary tool used for targeted queries for some forms, which each time had to be localized in the paper editions.

Putting together a collection of printed MA literary texts has not been an easy task. There is no bibliography of such works that one could rely on. They are not easily available either. The choice of modern titles written in MA in big and small bookstores in Rabat and Casablanca was limited to three in 2012 and it decreased to two upon my return two years later (without counting some collections of poetry). Some titles are hard to find in Moroccan state libraries. In this respect, my visits to the Library of the King Abdul-Aziz Al Saoud Foundation for Islamic Studies and Human Sciences in Casablanca were most rewarding. However, since my practical access to its resources was restricted by the library's rules prohibiting any photographing and allowing only a part of every book to be photocopied, some books are only partly analyzed.

I. Written Moroccan Arabic and Written Standard Arabic

The term ‘Moroccan Arabic’ should not suggest that there is one homogeneous Arabic dialect used in Morocco. Instead, it is normally used to refer to a group of varieties of Arabic spoken in this country. The area of what is nowadays Morocco was Arabized in two stages: the Islamic conquest in the 7th century and the conquest by the tribe of Banū Hilāl in the 12th century. These two migration waves are reflected in the two strata of dialects: pre-Hilalian and Hilalian. Heath (2002: 1-12) classifies Moroccan Arabic dialects into four major groups (the list below is based on a summary in Aguadé 2008: 288):

> (i) the northern type (sedentary, pre-Hilalian): Tangier, Tetouan, Jbala, ancient medina of Rabat, Fes, Sefrou, Taza, etc.,
> (ii) the Saharan type (Hilalian, Bedouin but now sedentary): Ḥassāniyya in Saharan towns like Mḥamid, Tata, Goulimine, Bedouin dialects (e.g. Zʕir) in the plains between Rabat and Casablanca,
> (iii) the central type (Hilalian, sedentary, rural): dialects of Oujda, Atlantic coast south of Rabat, Casablanca, rural dialects around Fes and Sidi Kasem, Atlantic coast south of Casablanca, El Jadida, Essaouira, Marrakesh, Skura, etc.,
> (iv) the Jewish dialects (pre-Hilalian, sedentary): in the traditional Moroccan towns, now almost completely disappeared.

This polydialectal situation is witnessing a process of koineization. The modern koine, based on the dialect of Casablanca, has developed as a result of the increasing influence of this commercial, industrial and political metropolis to which speakers of rural dialects are attracted in great numbers. This variety, most widespread in Morocco, is “structurally closest to the «central» type in its basic phonology (…) and grammar” (Heath 2002: 10). By way of simplification, it is this koine that is often referred to as the ‘Moroccan Arabic’.

As elsewhere in the Arab world, MA exists in the sociolinguistic situation of diglossia (Ferguson 1959 is the seminal paper on this concept). Diglossia means that a linguistic community uses two different but genetically related varieties of a language, usually characterizable as high and low, the functions of which are mostly complementary. The use of MA, the low variety with no codified standard for its spoken and written forms, is limited to determined, primarily oral, communicative situations, such as private conversation, popular music, popular

poetry and films of a secular character as well as written texts intended to reflect orality, e.g. theatre plays, comic strips or some dialogue parts in prose texts. This variety is acquired naturally by children by exposure in their domestic environment. Characterized as being of low prestige, it is sometimes considered by its very users to be a corrupted version of the real and pure language[12]. By contrast, the high variety, SA (Literary Arabic, *al-ʕarabiyya al-fuṣḥā*) has a codified standard for its both spoken and written forms[13]. Normally not resorted to in ordinary conversation, it is generally used in official situations such as public political and religious speeches, the press, television and radio news, books, including *belles lettres*, as well as science and education. The SA variety has no native speakers, but is acquired, to various degrees, sometimes on a very poor or passive level, by way of official instruction in schools[14].

---

[12] On the other hand, it seems that Arabs have a very intimate relation to their dialects, which is usually ignored or underestimated in the sociolinguistic descriptions. Arabs tend to see in their dialects the repository of their inner, private and social life, which should not be exposed to outsiders. These should be rather impressed with the richness, intricacy and deep-rootedness of SA.

[13] To be precise, being used in an extended area and having no pan-Arab authoritative regulatory organization to unify it, SA cannot be absolutely homogeneous. It suffices to have a look into one of its standard dictionaries, for instance Wehr (1980, and other editions), to find words or graphies the use of which is limited to particular countries and regions. As observed by Diem (2006: 2), SA is chiefly standardized as far as writing and morphology are concerned, whereas pronunciation and vocabulary are characterized by regional peculiarities. He also adds that the reason for the impression of the uniformity of SA is the confusing multitude and diversity of Arabic dialects.

[14] This dichotomous model has been found to be insufficient to properly account for the language situation in the Arabophone world. Descriptions have been proposed using the concept of multiglossia (polyglossia), i.e. a continuum of particular varieties or levels between which speakers can shift during one conversation (e.g. Badawī 1973; El-Hassan 1978; Meiseles 1980; Kaye 1994). As Bassiouney (2009: 15) observes, since such levels have no clear boundaries, "one could theoretically propose an infinite number of levels". Alosh (2005: 7) decisively argues against the "compartmentalization of styles into discrete language varieties".

Similar insights have also been made for the sociolinguistic situation of Arabic in Morocco. It has been described with the use of such concepts as triglossia, in which SA is differentiated from CA, and quadriglossia, in which, in addition, MA is differentiated from Educated MA (Ennaji 2005: 47-50). According to Youssi (1992, 1995), the Moroccan triglossic model comprises: (1) *arabe marocain* (Moroccan Arabic, i.e. 'Moroccan dialectal varieties in the daily and intimate usage"), (2) *arabe marocain moderne* (Modern Moroccan Arabic, also referred to as the middle variety [*arabe moyen*], i.e. 'the formal pan-Moroccan variety, intermediate between l'AM [arabe marocain], and l'AL [l'arabe littéral]'), and (3) *arabe littéral* (Literary Arabic, i.e. 'the ancient and modern varieties of written Arabic, irrespective of the period') (Youssi 1992: 16-17). Ennaji emphasizes that the sociolinguistic complexity of Morocco makes this country different from the rest of the Arab world (2005: 6-7), with Berber covering "the domains of home and street" together with MA and with French having "functions and domains which overlap with those of Standard Arabic, in addition to covering the private sector, science, and technology" (2005: 2). Bilingualism or multilingualism are other important linguistic characteristics of Moroccan society.

The primary domain of SA is writing but it also functions in speech (television, radio, public communication, education). By contrast, until recently, most Arabic dialects had functioned almost exclusively in speech. Literature written in dialect had been considered to be of low grade, since the only 'true' language of literature was SA, close to Classical Arabic, in which the Quran was revealed and classical and canonical texts were written. The complementarity balance, however, seems to have been tipped in favour of SA. As Bauer observes, nowadays, more people have knowledge of SA than before and it is used in speech on an everyday basis, while written dialect is still by far an exception (Bauer 1996: 1488, cf. also Diem 2006: 87-95 and Langone 2008). Nevertheless, in some parts of the Arab world, notably in Egypt and Lebanon, the dialect has begun to be perceived by some groups and individuals as an exponent of national, or at least regional, identity. To a lesser extent, this is also visible in Morocco, where since the 2000s[15], MA has had its champions on the literary, artistic, social, academic and journalistic levels. The use and empowering of MA is outspokenly advocated by figures such as the linguist ʕAbd al-Raḥīm Yūsī [Abderrahim Youssi] (e.g. Youssi 2013), author of a grammar of MA (Youssi 1992)[16], and writers Fouad Laroui (Laroui 2011) and Murād ʕAlamī, author of a number of literary translations into MA and a "novel in the vernacular Moroccan language" (included in the corpus as *Ṛḥ.*). MA has also received some legal acknowledgement: Article 5 of the Moroccan Constitution of 2011 stipulates that the State acts to protect "the dialects and cultural expressions" (*al-lahağāt wa-t-taʕbīrāt aṯ-ṯaqāfiyya*) used in Morocco.

The change of attitudes towards dialect and new communication technologies such as the Internet (social networking services, blogs, forums, etc.), SMS messages and others have enabled the vernacular varieties to be used in everyday informal electronic communication in written form, in the Arabic or Latin script (for Arabic in general, cf. Belnap & Bishop 2003; for MA: Berjaoui 2001; Benítez Fernández 2003; Caubet 2004, 2008: 274, 2012: 388-399, 2013, 2017 and Srhir 2016: 113-115). It is also used by some media, both printed and electronic (for MA see Aguadé 2012; Benítez Fernández 2012a and b; Miller 2012, 2017; Caubet 2012; Elinson 2013: 717-719; Hoogland 2013b, 2018; Langone 2003 and Srhir 2016: 110-113). As for literary creation, the dialect has started to be used in this

[15] Cf. Miller (2017: 91), who observes, in addition: "it is only in the late 1980s, early 1990s that the first public stands toward the valorization of dārija started to emerge" (p. 93). Elinson (2017: 194) places the beginning of the discussion around the topic of writing in MA as far back as in the mid-1970s.

[16] To be precise, it is a grammar of what the author refers to as Modern Moroccan Arabic (*arabe marocain moderne*), which is a form intermediate (*moyen*) between various Moroccan Arabic vernaculars and SA. As Durand (2004: 40) points out, this book is a historic work for more than one reason: it is of considerable scope (485 pages), was written by a Muslim native speaker of the described variety and was published in Morocco.

function on a larger scale, not only in poetry and theatre plays, but also in prose works[17]. This does not apply to some Arab countries, for instance, to Algeria, where it is nearly impossible to publish works in the dialect (Langone 2008: 54), or Syria, where the only literary work in Syrian Arabic I was able to get hold of in the numerous bookshops of Damascus in 2003 and 2004 (and I would be very surprised if the situation should have changed a lot by now) was the transcript of a popular radio drama, thus, a text primarily destined for oral performance. There are, however, countries in which more and more texts written in dialect are published every year. Egypt, with EA, the indisputable leader as far as written literary production in dialect is concerned (see Rosenbaum 2004 on this dialect gaining the status of a language of written literature and Elinson 2013: 727, footnote 4, for bibliography), is followed by Lebanon in this respect. As for modern literary creation in MA, it is considerably less rich[18]. Its use in contemporary *belles lettres*, including literary translations (see Elinson 2013: 723-726), has been described from a broader perspective by Ferrando (2012a), Elinson (2013: 719-726), Miller (2015, 2017) and Srhir (2016: 103-110), while studies devoted to one or more specific cases are: Aguadé (2005, 2008: 295, 2013), Ferrando (2012b), Langone (2006) and Hoogland (2013b). Let us remark at this point that despite the growing activity in the sphere of literary creation in MA, books of this kind are rarely available in bookshops and many of them are confined to library shelves.

I.1. Written Standard Arabic

The relation high variety–low variety (standard language–dialect) which binds SA and MA in terms of functions and prestige is reflected in another one: that of donor–recipient as far as script is concerned[19]. Since SA is the prestigious variety with a long writing tradition, it is natural for MA to borrow from it its graphic

---

[17] As Cachia observes, texts "of artistic quality" written in dialect "have long existed, even outside folk literature"; what was problematic was their acceptance within the literary canon (Cachia 2010: 190). On the use of Arabic dialects, in general, in literary production see Diem (1974: 96-125), Grotzfeld (1982), Bauer (1996: 1488-1490), Davies (2006), Walther (2010: 189) and Cachia (2010).

[18] Colin (1986: 1207) outlines the oldest poetical tradition in MA, the earliest texts being specimens of the poetry of the towns preserved in Ibn Khaldūn's *al-Muqaddima*, poems in honour of the Prophet, and *zağal* (all in the Spanish Arabic dialect). The dialectal poetry in the period that followed was referred to as *məlḥūn* (see Pellat 1991), the first known and most famous *qaṣīda* in *məlḥūn* being that by al-Maġrāwī (16th c.). For the contemporary situation of *zağal* see Elinson (2017).

[19] Arabic script has been adapted, for a larger or shorter period of time, for many other languages, the most notable cases being Persian, Turkish and Urdu (see Daniels 2013: 425 for a non-exhaustive list). The situation of MA, like other Arabic dialects, is special in this respect since in this case both the donor and the recipient are closely related genetically and structurally.

system and adapt it for its own needs[20]. It is not necessary to present the historical background of this script here[21], yet of importance is to mention the existence of two orthographic traditions, which, following Mumin & Versteegh (2014: 6-7), will be referred to as the Ḥafṣ tradition and the Warš tradition. They are associated with the two major Quran readers and exegetists, ʔAbū ʕAmr Ḥafṣ ibn Sulaymān ibn al-Muġīra al-Kūfī (d. 796) and ʕUṯmān ibn Saʕīd Warš al-Qurašī al-Miṣrī (d. 812). The Ḥafṣ tradition became dominant in the East of the Arab world, while the Warš tradition established itself in its Western parts. In addition to the issues of correct pronunciation and interpretation of the Quran, they differed in the form of the graphs marking three sounds: *f*, *q* and *n* (Mumin & Versteegh 2014: 6-7). Remnants of this divergence can still be seen today: While the prevalent SA graphy of *f* and *q* in the entire Arabophone world is ف ‹f› and ق ‹q›, respectively, in some texts printed in Morocco, the graphs with a different pointing (*naqṭ*), ڢ ‹f̣› and ڧ ‹f_q›, are used (see Sect. III.1.)[22]. The fact that these two sounds were written differently by people in the West (*ʔahl al-Maġrib*) and in the East (*ʔahl al-Mašriq*) was well known to the classical Arabic philologists concerned with the art of writing, both in the West (e.g. al-Dānī 1997: 37) and in the East (e.g. al-Qalqašandī [n.d.]: 154). Other significant geographical differences concern the notation of the word-medial glottal stop by means of the *hamza* sign (see Parkinson 1990), the use of the so-called Egyptian *yāʔ*, i.e. marking word-final *ī* by means of ى ‹á›, typically marking *ā*[23], or the notation of foreign sounds, particularly *g* (cf. Badawi *et al.* 2004: 17-19). This divergence results, to a considerable degree, from the

[20] The Latin script, with digits adapted for marking some sounds specific to Arabic, is another solution chosen by some native users of Arabic (Caubet 2004, 2012, 2013; Moscoso 2009; Benítez Fernández 2012a). One can thus speak of a digraphy, albeit not acknowledged or standardized, in the dialectal sphere of the Arabophone world. The informal Latin-based writing system for Arabic has no established name, but usually it is referred to as *al-ʕarabīzī* (a blend of *ʕarabī* 'Arabic' and *ʔinglīzī* 'English'), *ʕarabiyyat ad-dardaša* ('the chatting Arabic') or *al-frānkū*. Ubiquitous in the Internet and mobile phones, this phenomenon is especially interesting insofar as it is not always used as an alphabetical system for representing the phonetic form of spoken utterances but often retains the Semitic spelling principles of not marking short vowels or assimilations, e.g. that of the definite article *al-*. In this respect, this system is sometimes very close to the transliteration of Arabic script into Latin characters.

[21] For the origin and development of Arabic script see Gruendler (2006) and Daniels (2013: 420-424, 2014). For the development of SA orthography, see Diem (1976, 1979-1983, 1982), Goldenberg (2013: 38-39) and Saiegh-Haddad & Henkin-Roitfarb (2014: 15-19).

[22] In addition, a specific calligraphic style, *al-maghribi*, originated in the Western parts of the Arab world (see Houdas 1886; van den Boogert 1989; ʔAfā & al-Maġrāwī 2007 and Gacek 2008), but it has no direct relevance for the matters we are concerned with in the present work.

[23] The double phonetic function of ى ‹á› is reflected in the lack of unanimity among Arab philologists on how this graph should be referred to. Some authors (e.g. Qabbiš 1984: 103, 106) call it "*yāʔ* without dots", while others (e.g. al-ʔAsmar 1988: 63) consider it a type of the *ʔalif*, viz. the so-called *ʔalif maqṣūra* ('shortened *ʔalif*'), which has "the form of *yāʔ* without dots".

lack of an efficient regulatory body whose orthographic authority would extend over the entire Arab world. The national Arabic language academies in Damascus, Cairo, Baghdad, Amman, Tunis and Rabat exert only regional influence, while the Union of the Arabic Language Academies, established in 1971, is concerned mainly with standardizing terminology (Sawaie 2007: 640). The SA orthographic norms, usually referred to as *(qawāʕid) al-ʔimlāʔ* (from *ʔamlā* 'to dictate'), are a frequent topic of didactic publications, e.g. Qabbiš (1984), al-ʔAsmar (1988), Hārūn (1993), al-Ṭabbāʕ (1993) and Ḥasanayn & Šaḥāta (1998), and are sometimes included in manuals of Arabic grammar, e.g. al-Ġalāyīnī (2002 [1912]: 265-283) and Daḥdāḥ (1989: 12-14). There have been various proposals for reforming Arabic script, accounts of which are Al-Toma (1961), Meynet (1971) and Hamzaoui (1975).

I.2. An outline of Standard Arabic graphy

The SA alphabet (*al-ʔalifbāʔ*) is an inventory of 28 letters (primary graphs, see Sect. III.1. for this term). The Arabic script is not an alphabetic system but an abjad, i.e. a system in which only consonants are represented. Strictly speaking, it is an imperfect abjad because, in most cases, long vowels, too, are represented by means of letters (called, in this function, *matres lectionis*)[24].

Most graphs of the SA alphabet connect with the following graphs and are referred to as connectors. The non-connectors, which do not show this feature, are: the *ʔalif*, whether bare, ا ‹a›, or with a *hamza*, أ ‹å›, إ ‹ą›, or *madda*, آ ‹ã›, as well as د ‹d›, ذ ‹ḏ›, ض ‹ḍ›, ر ‹r›, ز ‹z›, و ‹w›, ؤ ‹ẘ› and, finally, the *hamza* sign ء ‹o› with no support, traditionally not counted as a letter of the alphabet (one of its oddities being that it is never connected on either side). There is one obligatory ligature, i.e. special shape for two connected graphs: لا for ل and ا (instead of لـا)[25]. Other ligatures are facultative and depend on the handwriting or font used.

The shapes of the graphs of the Arabic script are modified depending on their graphic environment, i.e. whether they are or are not connected to the preceding and following graph, or, as Coulmas (2003: 123) puts it, depending on their position in a writing group (see footnote 2). For instance, the four formally different units: غْ, غـ, ـغ and ـغـ all represent one sound, *ġ* in SA, and belong to one graph: غ ‹ġ›. It is superfluous to demonstrate that they belong to one set since it is known by convention. The formal differences between them are determined purely by linear-distributional factors (cf. the case of the Greek *sigma*, with two graphies:

[24] As Daniels (2013: 415) remarks, a pure abjad was used only for writing Phoenician.

[25] This ligature has sometimes been counted as a 29th letter of the alphabet (Daniels 2013: 413).

the non-word-final *σ* and the word-final *ς*), therefore they are sometimes described as positional variants of one grapheme (Polański 2003c: 38). This issue is of no special theoretical interest, but it will reappear in the context of the positional neutralization of heterography in Sect. III.1. Some graphs are not allowed to occur in certain positions and are obligatorily replaced. For instance, ى ‹á›, used primarily to mark *ā* in SA, occurs only word-finally. If any graphic unit is attached finally, this graph is obligatorily replaced by ا ‹a›. Likewise, under same circumstances, ة ‹ħ› is obligatorily replaced by ت ‹t›. Being systematically predictable and obvious, such situations are of no further interest to us here.

The spelling of *ʔ*, the glottal stop, is governed by special historically motivated rules. This sound is marked by means of ء ‹o›, i.e. the symbol of the *hamza*[26] written directly on the line (traditionally not included in the alphabet, but considered a graph on its own in this study), or its combination with another letter serving as its support. These letters are: ا ‹a›, i.e. the *ʔalif ṭawīla* (the *hamza* can be written above or below it), ى ‹á›, i.e. *yāʔ* without dots (called *nabra*), or و ‹w›. The resulting combinations: أ ‹å›, إ ‹ę›, ئ ‹ŷ› and ؤ ‹ẘ›, are considered separate graphs in this work. Which one of these graphs is used depends on the position of the sound within a word and its phonetic environment, but also other factors. One of them is, for instance, whether the word-medial position of the letter to mark *ʔ* is accidentally medial (*mutawassiṭ ʕaraḍan*, al-ʔAsmar 1988: 180; termed also 'quasi-medial', *šibh mutawassiṭ*, by al-Ġalāyīnī 2002 [1912]: 273), i.e. caused by attaching an affix to a graphic word ending in a *hamza* (*ʔ* being the final radical of the word), or if this position is inherently medial (in al-Ġalāyīnī's terms: 'properly medial', *mutawassiṭ ḥaqīqatan*, p. 273), i.e. the graph marking *ʔ* being a non-final radical of the word. To illustrate the graphy of *ʔ* in an accidentally medial position: *yaqraʔu* 'he reads' has only one spelling: يقرأ ‹yqrå›, i.e. *ʔ* is marked as أ ‹å›, its position being word-final. But *yaqraʔūna* 'they$^{m\ pl}$ read', in which the position of the graph to mark *ʔ* becomes word-medial because of adding the plural suffix *-ūna*, can be written (1) يقرؤون ‹yqrẘwn›, with *ʔ* marked as ؤ ‹ẘ›, i.e. in accordance with the rules for its word-medial position, or (2) يقرأون ‹yqråwn›, with *ʔ* marked as أ ‹å›, i.e. as if its position is still word-final (al-ʔAsmar 1988: 180-181). In al-ʔAsmar (1988: 129-193), a popular spelling book, more than thirty pages are devoted to the issue of marking this sound (cf. also Badawi *et al*. 2004: 11-14 and Parkinson 1990 for an analysis of inconsistencies in its spelling).

Some SA phonetic elements are marked in writing by means of additional signs termed vocalization or, less often, vowelization (*taškīl*). Vocalization signs constitute a secondary graphic system, which, being unable to exist on its own, is

[26] In Arabic, this term refers both to the sound and the graphic symbol marking it.

superposed on the largely self-sufficient primary system of graphs marking consonants and long vowels[27]. The vocalization signs are used to mark:

a. short vowels and vowellessness (the signs being termed *ḥarakāt*, see 'Transliteration and transcription' for details),
b. the nominal affix *-n* (the sign being referred to as *tanwīn*, or 'nunation' in English, from *nūn*, the name of the letter ن ‹n› marking this consonant),
c. the lengthening (gemination) of consonants, elision and the sequence *ʔā* (the signs being termed *šadda*, *waṣla* and *madda*, respectively, known collectively as *ḍawābiṭ*).

The use of these signs is facultative because the phonetic units they represent can nearly always be inferred from the context by readers having a sufficient command of Arabic (cf. the principle of economy in writing, below). The real usage of vocalization can be described as a continuum with one extreme constituted by fully vocalized texts, e.g. editions of sacred books (the Quran and the Holy Script) and critical editions of classical texts, and the other by completely unvocalized texts, e.g. a piece of news in a newspaper. In an average text, the general tendency is to keep it close to no vocalization, the *šadda* being most likely to be used, and to mark vowels in ambiguous contexts, i.e. if the two (or more) different pronunciations and meanings result in semantically coherent utterances. One single sign may suffice to indicate the intended reading, as in (1), showing a news heading from an online news service:

(1) أزمة العَلم تتصاعد بين بغداد وإقليم كردستان

‹åzmħ_**alʕ^{a}lm**_ttṣaʕd_byn_bġdad_wąqlym_krdstan›[28]
ʔazmat^{u} ***l-ʕalam^{i}*** *tataṣāʕad^{u} bayna Baġdād wa-ʔiqlīm^{i} Kurdistān*
'The crisis of **flag** escalates between Baghdad and Kurdistan'.

Without the *fatḥa* - ‹^{a}› marking *a*, the graphic word marked in bold would have the form العلم ‹alʕlm›, and thus would be ambiguous since another reading: *al-ʕilm^{i}* 'of science' would be possible. As a matter of fact, the problem of semantic ambiguity is relative and in most cases can or cannot be solved *pro captu lectoris*.

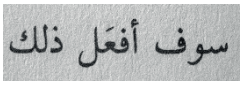
The use of vocalization signs can be quite unpredictable and irregular. Sometimes, they may be not written when they would be helpful, while at other times,

---

[27] One could also distinguish tertiary graphic signs, i.e. indications for reciting the Quran, and perhaps even quaternary elements: ornamental fillings in Arabic calligraphy.

[28] http://www.aljazeera.net/news/arabic/2017/4/2/أزمة-العلم-تتصاعد-بين-بغداد-وإقليم-كردستان (accessed 5 Dec 2017).

more often, they are written when they are superfluous. The graphic word which a vocalization sign is supposed to clarify may be clear even for a first-year student of Arabic, as exemplified in (2).

(2) 

‹swf_åfʕ^{a}l_ḏlk› (*Ṯaʕ.* 26[3])
sawfa ʔafʕal^{u} ḏālika
‘I will do that’

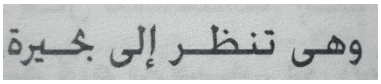
The *fatḥa* ‹^{a}› is absolutely redundant here as the unvocalized graphic word ‹åfʕl› can have no other reading than *ʔafʕal^{u}* ‘I do’ in this context.

In classical writing, it was a common practice to signal the exact phonetic form of a word by means of the *ḍabṭ*, i.e. describing in words whether a given letter is dotless (*muhmala*) or has so and so many dots (e.g. *muṯallaṯa* ‘tripled’, i.e. triply dotted), whether it is followed by *a* (*maftūḥa* or *bi-l-fatḥ*), *u* (*maḍmūma* or *bi-ḍ-ḍamm*), *i* (*maksūra* or *bi-l-kasr*) or no vowel at all (*maskūna* or *bi-t-taskīn*). Instances of this procedure can be occasionally encountered in contemporary usage as well, for example in the following phrase:

(3) أناديها حبيبتي، بتسكين الحاء
‹ånadyha_ḥbybty_btskyn_alḥao› (*FB* 66$_{5\text{-}6}$)
ʔunādī-hā ḥbibt-i bi-taskīni l-ḥāʔ
‘I call her *ḥbibti* (my love), with a vowelless *ḥ*’

The *ḍabt* is used here, instead of a *sukūn*, to indicate that the graphic word حبيبتي ‹ḥbybty› represents the dialectal *ḥbibti* ‘my love’, not SA *ḥabībatī*, although it is used in a SA context.

The redundancy of vocalization is strictly linked to an important feature of the Arabic script: context-dependence, i.e. the role of context in solving representational and semantic ambiguities. Only exceptionally does this context-dependence concern letters, i.e. the primary signs. This is, for instance, the case of the following graphic expression:

(4) 

‹whá_tnẓr_ạlá_bḥyrħ› (*Taġ.* 245[7])
wa-hiya tanẓur^{u} ʔilā buḥayra^{tin} ‘and she is looking at a lake’

or

wa-hiya tanẓur^{u} ʔilayya bi-ḥayra^{tin} ‘and she is looking at me with perplexity’

The phrase is taken from a book published in Egypt, where the graph ى ‹á›, apart from marking the final *ā*, is also used to mark the final *ī* or *y*. Because of its phonetic ambiguity, the graphic word which contains it is ambiguous, too. It may represent either: *ʔilā* ‘at, to’ or *ʔilayya* ‘at me’ (each reading entails a different reading of the graphic word that follows: *buḥayra^{tin}* ‘a lake’ or *bi-ḥayra^{tin}* ‘with perplexity’, respectively). However, the context (there is no lake to be seen and the first-person narrator speaks of a woman talking to him) allows only the latter interpretation.

Variational heterography, i.e. the use of various graphies for the same phonetic word, can be encountered in SA, although its standardization makes this phenomenon seem rather infrequent, especially in comparison with the graphy of Arabic dialects. Linear variational heterography in SA is rather rare[29]. It concerns, e.g., the compound conjunction *kay-lā* ‘in order not to’, which can be spelt كيلا ‹kyla› or كي لا ‹ky_la› (e.g. al-ʔAsmar 1988: 282). More instances are found if non-canonical graphies are taken into consideration. For instance, the orthographic rules require the auxiliary *mā zāla* ‘to keep on, not to cease’ to be spelt ما زال ‹ma_zal›, but the graphy مازال ‹mazal› is very often used. Contrary to classical rules, teophoric names, for instance *ʕAbd al-Karīm*, lit. ‘Servant of the Generous’, have frequently continuous graphy, viz. عبدالكريم ‹ʕbdalkrym›, instead of being written as two graphic words: عبد الكريم ‹ʕbd_alkrym›. Quantitative variational heterography affects mostly loanwords and foreign proper names, e.g. *film* ‘film, movie’ is spelt as both فيلم ‹fylm› and فلم ‹flm›, but also some native Arabic words, e.g. *miʔa* ‘hundred’ is written مائة ‹maŷħ› and مئة ‹mŷħ›[30]. Qualitative variational heterography in SA can often be observed in marking foreign sounds such as *g*, e.g. *garāž* ‘garage’ is usually spelt جراج ‹ğrağ› in Egypt, but كراج ‹krağ› in Syria. It also affects native Arabic words: Thus, the final *ī* or *y* has two graphies ي ‹y› and ى ‹á›, e.g. *ʔilay-ya* ‘to me’ can be written إلي ‹ạly› or إلى ‹ạlá›. The variational spelling of *ʔ* was mentioned some paragraphs earlier. Some of these phenomena, including the qualitative graphy, are mirrored in the graphy of MA, and will be discussed in respective paragraphs.

---

[29] Discontinuous and continuous graphy are described in traditional Arabic grammar under the headings *faṣl* ‘separating’ and *waṣl* ‘joining’, respectively (e.g. al-ʔAsmar 1988: 279-299).

[30] In traditional Arabic grammar, the issue of quantitative variational heterography is described in terms of two procedures: *ziyāda* ‘addition, augmentation’ and *ḥaḏf* ‘deletion’ (e.g. al-ʔAsmar 1988: 247-276), the latter also called also *naqṣ* or *ʔinqāṣ* ‘diminishing’ (al-Ṭabbāʕ 1993: 110). The former consists in adding a letter that marks no sound, while the latter means that a sound is not marked by a letter. Thus, the point of reference for both procedures is the phonetic form. In the present study (Sect. III.2.1.2.), no differentiation of this kind is made.

I.3. Orthographic principles in Standard Arabic graphy

Some orthographic principles identified by scholars, mostly for Latin-based scripts used in Western languages, underly the spelling of SA as well. Different authors propose and discuss different sets of such principles[31], out of which the phonetic (not always distinguished from phonological or phonematic) and morphological principles appear most frequently (e.g. in Dürscheid 2006: 142, a university course book on the linguistics of writing). These principles should be conceived of in terms of mere tendencies, the interrelations of which are often unsystematic and the effects of which may overlap (cf. Bußmann 1990: 628). For a given spelling, more than one underlying principle can be identified, or, in other words, a spelling resulting from one principle may be coincident with the result of another. In what follows, orthographic principles underlying SA graphy are identified and discussed, partly with the use of the concepts given in Bußmann (1990: 628-629).

1. Phonetic principle: the spelling reflects the pronunciation of the word so that one sound corresponds to one written sign and, ideally, *vice versa*[32]. For such a principle to be applied, there must exist an inventory of relatively stable correspondences between graphs and sounds, usually in the form of an alphabet. The phonetic values of graphs are usually fixed only to some extent, e.g. in English, the graph ‹c› is used to mark more than one sound: [s], [k] and, less often, [ʧ]. In SA, the phonetic principle underlies the graphy د ‹d› in ازدهر ‹azdhr› *ʔizdahara* ‘he/it^{m} flourished’, in which it marks the sound *d*, although on the morphological level the transfix of this verb is *ʔi-ta-a-*, with *t*.

2. Phonological principle: one written sign corresponds to one phoneme. For example, the SA graph ا ‹a› marks [aː], [æː] or [ɑː], three sounds that belong to one phoneme (other functions of this graphs can be disregarded at this point). Another instance of phonological spelling concerns the feature of emphasis (manifested in pharyngealization or velarization), which is not graphically represented if a non-emphatic consonant becomes emphaticized due to its linear vicinity to an emphatic sound (emphasis spread) (cf. Saiegh-Haddad & Henkin-Roitfarb 2014: 23, esp. footnote 24).

---

[31] For instance, Bußmann (1990: 628-629) distinguishes nine for German, while Polański (2016b: 11-12) gives four for Polish.

[32] It goes without saying that phonetic spelling is different from phonetic transcription in the scientific, linguistic sense (in order to stress this, in Michalski 2017, the term “pronunciation-oriented” was used rather than “phonetic”).

3. Morphological principle: the spelling represents the abstract morphological structure of a word without reflecting phonological processes such as voicing, devoicing, assimilation, etc. In other words, the graphy reflects the phonetic form of the basic (unmarked) allomorph. This principle is pervasive in SA graphy. For instance, *ʔaradtu* 'I wanted', pronounced [ʔaratːu], is written أردت ‹årdt›, with د ‹d›, because its abstract root is *r-w-d* (not **r-w-t*) and the sound *d* is pronounced when no factors cause devoicing, e.g. in *ʔarāda* 'he wanted', pronounced [ʔaraːda]. The morphological principle enables words belonging to one inflectional paradigm to have a common spelling of their lexical morpheme despite different phonetic forms. This principle is also responsible for the definite article in words beginning with assimilating consonants being spelt as ‹al› although the phonetic form contains a lengthened consonant, not [l]. For instance, [ʔazːajt] 'oil^{def}' is written الزيت ‹alzayt› (not *ازّيت ‹azːayt›)[33].

4. Historical principle is responsible for preserving relics from earlier stages of the language in the graphy of some words. As Polański (2016b: 12) puts it, it underlies spellings which cannot be accounted for by contemporary language awareness, but reflects historical processes of the development of a language. It manifests itself in the graphy of numerous SA words. The most prominent cases are:

a. The graph ا ‹a› is written facultatively in the graphic word مائة ‹maẘħ› *miʔa^{tun}* 'hundred'. It is said to have been introduced as a disambiguating element at a time when dotting was not used and the graphic word for 'hundred' could be confused with *min-hu* 'from him' (al-Ġalāyīnī 2002 [1912]: 268; Qabbiš 1984: 76)[34]. Nowadays, with the two meanings clearly differentiated in writing even without ا ‹a›, cf. مئة ‹mẘħ› for *miʔa^{tun}* and منه ‹mnh› for *min-hu*, this graph has become redundant.
b. The *ʔalif al-wiqāya*, lit. '*ʔalif* of prevention' (known also as the *ʔalif al-fāriqa* 'the separating ʔalif' or *ʔalif al-fāṣila* 'idem', sometimes referred to as the 'otiose *ʔalif*') is written word-finally in the 2nd and 3rd person personal plural past and non-indicative present tense forms of the verb, e.g. in كتبوا ‹ktbwa› for *katabu* 'they wrote'. Its function was to distinguish و ‹w› marking the verbal suffix *-ū* from و ‹w› marking the conjunction *wa-*

[33] The morphological principle corresponds to Bußmann's etymological principle (1990: 629), according to which etymologically related words are spelled analogously.

[34] Not very convincingly, Wright (1974 [1859-1862]: 258) attributes the use of ا ‹a› to "a piece of bungling" on the part of the oldeśt writers of the Quran, who probably thought it to indicate the sound *a*, but inadvertently misplaced it. Let us also remark that this archaic graphy is taken at face value by some speakers of SA, including television newsreaders, who mispronounce the word for 'hundred' as [mæːʔa] insťead of [miʔa]. In this way, the redundant *ʔalif* acquires phonetic value.

‘and’ at a time when no spaces were used to separate words (Wright 1974 [1859-1862]: I, 11; Holes 2004: 92). According to another explanation, given e.g. by Qabbiš (1984: 77), it distinguishes verbs from nouns on the one hand and radical units (*ʔaṣliyya*) from inflectional ones (*ṭāriʔa*) on the other hand. Nowadays, with spacing used consistently in typography, it has turned redundant but remains obligatory.

c. There are various graphies of the *hamza* symbol used to mark *ʔ*. The reason for this is that while the original spelling was based on West Arabian Ḥiğāzi forms, e.g. *suwāl* ‘question’, written سوال ‹swal›, the classical phonetic standard subsequently adopted relied on East Arabian pronunciation: *suʔāl*. Eventually, the *hamza* sign was superimposed on the original spelling (Goldenberg 2013: 39; cf. Beeston 1970: 26), resulting in سؤال ‹sŵal›, a graphy which is in use currently. Modern presentations of the Arabic spelling explain the rules for marking *ʔ* differently: in terms of the phonetic environment of this sound.

d. The graph ى ‹á›, i.e. the *ʔalif maqṣūra*, marks final *ā* in such verbs as نمى ‹nmá› *namā* ‘he/it[m] grew’. Again, the spelling ى ‹á› reflects West Arabian pronunciation *-ay* (or similar), while the SA phonetic norm is based on East Arabian pronunciation *-ā* (Beeston 1970: 27). Nowadays, the current explanation of this graphy is synchronic and morphological: ى ‹á› is said to indicate that the third radical of the verb is *y*, not *w*. In the latter case, *ā* is marked as ا ‹a›, for instance in نما ‹nma›, which has the same pronunciation and meaning.

e. The current shape of ة ‹ħ›, called the *tāʔ marbūṭa*, lit. ‘the linked *tāʔ*’, reflects its two historical phonetic functions. First, it used to reflect the non-classical pronunciation *h* word-finally in all contexts (with *h* being marked as ه ‹h›, a graph from which ة ‹ħ› derives its basic shape, the so-called *rasm*), while the two dots, characteristic of ت ‹t›, were added to it in order to make it reflect classical pronunciation *t* (Beeston 1970: 27, cf. also Fischer 1972: 10 and Holes 2004: 92)[35]. However, nowadays, its functions in SA are different. First, it marks *t* in some nouns in the construct state (the first term of the genitive construction), e.g. in *laylatu* ‘night’ used in ليلة العيد ‹lylħ_alʕyd› *laylatu l-ʕīd*[i] ‘the night of the feast’. The second function is actualized in the pausal form of such nouns. The pausal form of the word for ‘night’, *layla*, is written in the same way. If the fully vocalized

[35] The transliteration symbol “ħ”, i.e. an “h” with a horizontal stroke, used in this study, is an attempt to reflect this formal two-facedness. Obviously, it should not be confused with [ħ] used as an IPA symbol.

form, لَيْلَةٌ ‹l^{a}y$\cdot$l^{a}ħ^{un}›[36] is considered, ة ‹ħ› does not represent any sound ("is silent", according to Wright 1974 [1859-1862]: I, 7, footnote). It seems, therefore, that ة ‹ħ› has been re-interpreted as marking the final *a*, taking over this function from the *fatḥa*, usually not written[37]. Finally, it should also be added that in nouns in the free state (i.e. those that are not qualified by a noun in the genitive), including those in the pausal form, ه ‹h› is sometimes used instead of ة ‹ħ› in less conscientious spelling.

5. Homonymy principle: words differing in meaning that have the same phonetic form (and usually morphological structure) are distinguished in spelling. The few examples that can be found in SA are proper names as opposed to (semantically related) common nouns. For instance, *yaḥyā* meaning 'he/it^{m} lives' is spelt يحيا ‹yḥya›, while the homophonous male given name *Yaḥyā* is written يحيى ‹yḥyá› (cf. Qabbiš 1984: 105).

6. Principle of economy: reduntant elements are not written. Its most notable manifestation is the practice of not marking short vowels, consonanst lengthening and other units (represented only occasionally as vocalization signs). In addition, in some words, even long vowels are not written: e.g. *ā* in the demonstratives such as *hāḏihi* 'this^{f}', written هذه ‹hḏh› (not *هاذه ‹haḏh›).

7. Principle of aesthetics: some graph combinations are avoided. A manifestation of this principle in SA is the avoidance of putting two identical letters together (*karāhat iğtimāʕ al-miṯlayn*). For instance, while *suflā* 'lower^{f}' is written سفلى ‹sflá›, with word-final ى ‹á›, i.e. the dotless *yāʔ*, the morphologically analogous *ʕulyā* 'upper^{f}' is written عليا ‹ʕlya› (not *عليى ‹ʕlyá›), in order to avoid the sequence of two *yāʔ*s, even though one would be dotted and the other dotless. Similarly, some spellers of SA prefer to write *ruʔūs^{un}* 'heads' as رءوس ‹rows› rather than رؤوس ‹rẘws›, in order to avoid two *wāw*s occurring together, even though one of them is merely a support of the *hamza*.

The above list of orthographic principles is still insufficient to account for all graphic phenomena in SA. For instance, graphs with no phonetic value are used to avoid the homography of different meanings and phonetic forms: Thus, the

[36] The nominative case and indefiniteness of this word, marked by ‹^{un}›, chosen arbitrarily for this example, are of no relevance here.

[37] This is reflected in the way the pronunciation of ة ‹h› is described in Weśtern scholarly literature. For inśtance, according to Beeśton (1968: 12), this letter "indicates a pronunciation which fluctuates between *-at* and *-a*". Similarly, Król (2009: 39-40) characterizes it as being pronounced in Modern Arabic as short *a* (in contexts other than the conśtruct śtate).

function of و ‹w› in عمرو ‹ʕmrw› is to ensure that this graphic word is interpreted as the name *ʕAmr*, not *ʕUmar* (the latter being written عمر ‹ʕmr›)[38]. According to another principle which was operative in Arabic writing from its beginning, the graphy of a word reflects its pronunciation in isolation (Goldenberg 2013: 38). For instance, the word-initial graph ا ‹a› marking the prothetic *ʔa* in *ʔaz-zayt*^{u} 'oil^{def}', written الزيت ‹alzyt›, is retained even if the prothesis is elided as a result of prefixing a unit ending in a vowel, such as the conjunction *wa-* 'and'. Thus, *wa-z-zayt*^{u} 'and oil^{def}' is written والزيت ‹walzyt›, with the graph ا ‹a› retained[39]. The final issue is the word-medial graphy of *ʔ*, which is different in the same phonetic environment depending on whether it is treated as being word-medial accidentally (e.g. as a result of prefixing a definite article to a graphic word with an initial *hamza*) or as being inherently word-medial (i.e. if *ʔ* is a non-final radical of the word). To illustrate this, the sound *ʔ* in the phonetic environment *-lʔu-* is marked as أ ‹å› in الأم ‹alåm› *ʔal-ʔumm*^{u} 'mother^{def}', but as ؤ ‹ẘ› in الؤم ‹alẘm› *ʔulʔum* 'be vile!'. The reason for this differentiation is that the former graphic word retains the word-initial graphy without the definite article (cf. أم ‹åm› *ʔumm*^{un} 'mother'), with the word-medial position of the graph marking *ʔ* being accidental.

MA is a special case as far as the principles underlying its graphy are concerned. Since it has borrowed its script from a genetically and structurally related donor language which is a prestigious variety, the donor-receiver relationship is an important factor shaping its graphy. In order to describe it adequately, yet other principles need to be proposed, in addition to those operating in SA. This will be done in Ch. V with respect to graphies affected by variational qualitative heterography.

---

[38] Since the graph ‹w› used in this case is a remainder of the Nabatean roots of the Arabic graphy (al-Ḥasan 2003: 111), it could also be interpreted as a manifestation of the historical principle.

[39] An exception to this rule concerns the elision of the prothetic *ʔa* in the definite article to which the preposition *li-* 'to, for' is attached. In this situation, the graph ا ‹a› is never written, e.g. *li-z-zayt*^{i} 'for oil^{def}' is spelt للزيت ‹llzyt›, not *لالزيت ‹lalzyt›.

II. General characteristics of Moroccan Arabic written in Arabic script

The issue of the Arabic script used for writing modern MA has been approached by scholars from various angles[40]. The articles by Aguadé (2005) and (2013) are case studies of MA graphy as used in particular books, while Aguadé (2006) is a comprehensive description of MA spelling practice, based on a heterogeneous selection of texts. Hoogland (2013a) analyzes a corpus of texts, belonging to various genres, in order to find common elements in MA graphy[41]. By formulating some recommendations, this study also represents the normative aspect in dealing with MA. Additional graphic phenomena in MA were described by Michalski (2016), who used an enlarged corpus of literary texts. Chapters or passages devoted to the graphy of MA are in Benítez Fernández (2010: 218-220) and, for MA used online, Caubet (2017: 133-136). Al-Midlāwī (2001) is a work of partly normative character as it contains proposals for marking in writing some specifically MA sounds.

II.1. Previous research on the qualitative features of Moroccan Arabic graphy

In the studies enumerated above, several important phenomena of various types: qualitative, quantitative and linear (see Sect. III.2.1. for a discussion of these concepts) have been identified in MA graphy. There is no need to recapitulate all of them here. The following presentation is limited to the most important issues related to qualitative variational heterography.

a. Non-Arabic sounds, *p*, *v* and *g*, are represented as ب ‹b›, ف ‹f› and ك ‹k›, respectively, in older texts due to the lack of special characters in printing presses, but since the generalization of the use of computers they have been marked as پ ‹p›, ڤ ‹v› and گ ‹ǵ› (Aguadé 2006: 255). The sound *g* is marked by means of ق ‹q›,

[40] For observations on the spelling of Arabic dialects in general see Holes (2004: 92-95), also minor remarks in Saiegh-Haddad & Henkin-Roitfarb (2014: 22). For a comprehensive discussion of the graphy of EA, see Rosenbaum (2004).

[41] His early work (1983) was unfortunately unavailable to me.

ك ‹k›, گ ‹ǵ›[42] and ڴ ‹ĝ›, as well as ج ‹ǧ› in words in which it corresponds to SA *ǧ* and in borrowings from Berber and European languages (in the latter case, exceptionally by means of چ ‹č›, Aguadé 2006: 259; Benítez Fernández 2010: 219-220). Rather marginally, it is marked by means of غ ‹ġ› (Michalski 2016: 386) and, in one source, as كپ, a digraph composed of ك ‹k› and "a *ya* (ى ‹á› – M.M.) with three dots below it" (Hoogland 2013a: 71)[43]. Thus, *g* is the sound with the by far largest number of evidenced graphies: as many as eight.

b. Letters marking interdental fricatives in SA are sometimes used to mark dental stops in MA if the latter correspond diachronically to the former (cf. Aguadé 2006: 257; Hoogland 2013a: 63, 72; Benítez Fernández 2010: 219). Michalski (2016: 385-386) observes a pseudo-correct use of these letters for marking sounds that are not interdental fricatives either in MA or SA, e.g. الذود ‹alḏwd› *d-dud* 'worms' (*BT* 116$_3$) (cf. SA الدود ‹aldwd› *ad-dūd*^{u} 'idem')[44].

c. Pharyngealized consonants resulting from secondary pharyngealization are marked by means of letters used for plain (non-pharyngealized) consonants. According to Aguadé (2006: 258), this is "normally" the case, whereas the use of letters marking pharyngealized consonants is rare (cf. also Benítez Fernández 2010: 219).

d. The suffixed personal pronoun of the 3rd person masculine singular *-u* 'him, his; it^m, its^m' is written as ه ‹h› or و ‹w›, the latter graphy being used more often according to Aguadé (2006: 262-263) and Hoogland (2013a: 69).

e. The prefix of the 3rd person non-feminine present tense form, *y-/i-* (cf. discussion under "ا ‹a› ~ إ ‹ą›" in Sect. VI.1.) is normally written ي ‹y›, but also إ ‹ą› (or إي ‹ąy›), as observed by Aguadé (2006: 265).

---

[42] In the enumeration given in the Arabic script in Aguadé (2006: 259), گ and گ are shown beside each other as two distinct graphs. I am not sure if these two actual graphs, differing only in the shape of the bottom stroke (a result of using two different fonts?), deserve being described as representing two different graphs. Examples adduced by the author suggest that he might have had گ and ك in mind. See also Sect. III.1. for the discussion of ڴ.

[43] Hoogland counts this "strange" graphy, which he recorded in *Ami.*, among "peculiar cases". I could not confirm the existence of this phenomenon since I had no access to this source. However, according to Mion (2014: 193), this sound is marked as گ ‹ǵ› there. The discrepancy between the two authors cannot be a result of them relying on two different editions of *Ami.* (2009 and 2014), because Mion (2014: 188) consulted both of them. A different question is whether the digraph described by Hoogland involves a *yāʔ* with three dots below it (ى) or rather پ ‹p›, since the difference between them is graphically neutralized in تكپلس *təgləs* 'you sit down', the example adduced by Hoogland. In either case, whether having پ ‹p› or ى ‹á› with three dots below as its second component, it would be the second digraph I have been able to identify in the MA graphy (the other one is تش ‹tš› used to mark *č*, see discussion under "GRAPH چ ‹č›").

[44] An analogous phenomenon in EA has been described by Rosenbaum (2004: 289-290).

f. The final vowel *a* is written in a number of ways. Aguadé (2013: 214) observes that in *Mi.*, ى ‹á› indicates that the 3rd radical of a verb is *y*, while ا ‹a› is used as a rule in other sources he examined. In addition, in *Mi.*, ة ‹ħ› is "naturally" used to mark the feminine gender of nouns and adjectives, although it also appears in the intensifying adverb *zəʕma*, spelt زعمة ‹zʕmħ›, while other adverbs are written with the final ا ‹a› there (2013: 214). Hoogland remarks that "different authors have made different choices" (2013: 70), but does not specify this with respect to word classes or grammatical categories. According to Michalski (2016: 388-390), who observes strong variation in the corpus which he used, there are as many as ten ways to mark the final *a* if vocalization signs and zero sign are taken into consideration: Apart from the usual ا ‹a›, ى ‹á›, ة ‹ħ›, ه ‹h›, ‹ᵃ› and the zero sign, there are also infrequent instances of أ ‹å›, اة ‹aħ›, اه ‹ah› and ي ‹y›. Verbs are spelt with either ا ‹a› or ى ‹á›, with "a considerable freedom", the choice between them being unrelated to the root of the verb. In the imperative, however, ا ‹a› is dominant. In nouns, the variation is between ا ‹a›, ى ‹á› and ة ‹ħ›, while in adverbs ه ‹h›, ‹ᵃ› and the zero sign are used, in addition.

g. The word-medial and word-final *a* in words that have SA cognates with the corresponding *ʔ* can have the phonetic spelling ا ‹a› or the "classicizing" graphy أ ‹å›. It was also observed that the abstract classical rules of the *hamza* spelling are sometimes applied to the phonetic reality of MA words (Michalski 2016: 388).

Major spelling principles (or tendencies) operating in MA have been observed in previous research, of which the two most conspicuous were characterized by Aguadé (2006: 255) as follows:

> (…) when writing in dialect Moroccans have two opposite possibilities: either to preserve as much as possible the orthography of Classical Arabic or to innovate trying to represent the phonemes of the spoken language: the result is generally a fluctuation between both tendencies[45].

In the present study, the former principle will be referred to as the *principle of donor-orientation*[46] (resulting in donor-oriented graphies[47]), while the latter corresponds to the phonetic principle, a special case of what is termed here the *principle of self-orientation* (resulting in self-oriented phonetic graphies).

---

[45] Such a situation was observed earlier for EA by Rosenbaum (2004: 284).

[46] This term intends to name the donor-receiver relationship between SA and MA in a language non-specific way. I also decided to employ it instead of the term 'classicizing' (used in Michalski 2016) in order to avoid the suggestion that what MA imitates are forms limited to CA, while in reality they are mostly SA forms.

[47] Aguadé calls them "*fuṣḥā* oriented graphies" (Aguadé 2006: 265).

The principle of donor-orientation is responsible for the graphy of any MA unit which cannot be interpreted as self-oriented (phonetic or morphological, discussed below) but is identical with the graphy used in its SA cognate[48]. For instance, the graphy ث ‹ṯ› used in MA ثاني ‹ṯany› *ṯani* 'second^{m}' to mark *ṯ* is neither phonetic or morphological. Instead, it copies the graphy of the SA cognate of this word: ثانٍ *ṯānin* 'idem'[49].

The principle of self-orientation divides into two subprinciples: phonetic and morphological. The *phonetic principle* means that the spelling of a given MA unit reflects its pronunciation. As mentioned in Sect. I.3., this requires a relatively stable correspondence between graphs and sounds, normally inventoried in an alphabet in which the graphs are given names indicating their basic, typical phonetic value[50]. For a language or dialect that has no alphabet of its own, like MA, positing the phonetic principle presupposes that it relies, more or less regularly, on the graph-sound correspondences obtaining in the alphabet of the donor language. For instance, in تاني ‹tany› *tani* 'second^{m}', the sound *t* is marked as ت ‹t› in accordance with the SA graph-sound correspondence. Being a more complex issue, the phonetic principle is elaborated upon in Ch. V.

The other subprinciple of the principle of self-orientation, underlying a considerable number of graphies, is the *morphological principle* (cf. Michalski 2016:

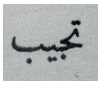
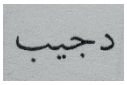
[48] For the needs of this study, two words are considered cognates if they share a common lexical and/or morphological structure. *Lexical cognates* share radical consonants and semantics, even though their morphological structures differ. For instance, MA *tqal* 'it^{m} became heavy' is structured following the morphological pattern $C_1C_2aC_3$, which differs considerably from $C_1aC_2uC_3$-, the pattern of its closest SA cognate, *ṯaqul-* 'idem'. In spite of this, these two words can be compared with respect to the graphs marking their radicals, e.g. *t* and *ṯ*. In lexical cognacy, the donor form-oriented graphy does not have to consist in transferring an entire SA form to a MA text without any changes; it may rather consist in preserving some particular features, like, in the present case, the graph marking a particular radical in the SA cognate.

Morphological cognacy, by contrast, binds MA and SA words having the same morphological patterns. This concept turns out to be of relevance if graphic variation affects an affix. For instance, the SA lexical cognate of MA *išuf* 'he looks' is SA *yatašawwaf^{u}* 'he expects', but it is morphologically too distant for comparison. In such cases, a morphological cognate is taken into consideration, for instance, SA *yašūb^{u}* 'he mixes, he corrupts', as far as the imperfect prefix is concerned (*i-* in MA, *ya-* in SA). Words that are both lexical and morphological cognates are *full cognates*, e.g. MA *iqul* or *igul* 'he says' and SA *yaqūl^{u}* 'idem'.

[49] The use of donor-oriented graphies can be understood in terms of bivalency, a phenomenon defined as "the simultaneous membership of a given linguistic segment in more than one linguistic system in a contact setting" (Woolard & Genovese 2007: 488). In the present case, bivalency consists in a given graphy being used in the graphic systems of both MA and SA.

[50] The reason for adding the qualifications 'basic' and 'typical' here is that some SA graphs are used in secondary functions to mark non-standard sounds (and these functions tend to be reflected in dialectal graphies). For instance, ج ‹ǧ› in Egyptian SA (and in EA), ق ‹q› in Saudi SA (and Saudi Arabic), ‹ǧ› and ‹k› in Levantine SA (and in Levantine Arabic) are all often used to mark *g*.

387, where it is shown to be responsible for some cases of quantitative heterography). It is illustrated in (1), showing two heterographs of the word *džib* 'you bring'. Form *A* represents the (frequently used) morphological graphy, with ت ‹t› marking the abstract morphological affix *t-* (assimilated to *d-* at the phonetic level), while Form *B* represents the phonetic graphy (used marginally).

(1)  

‹**t**ğyb› (*ṢX* 92$_8$) ‹**d**ğyb› (*Ṣa.* 253[1])
džib (< *t-ğib*) 'you bring'

A graphy which is identical with the SA graphy but can be interpreted as phonetic or morphological is said to be *phonetic* or *morphological* (respectively) *coincident with the donor graphy*. The graph ن ‹n› in ثاني ‹ṯany›, marking *n* phonetically, is an example thereof.

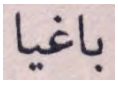
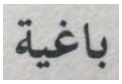
Both opposite principles, that of donor-orientation and that of self-orientation, can be operative in a single graphic word (cf. also Aguadé 2006: 258). For instance, in راس ‹ras› *ṛaṣ* 'head', the graphy ا ‹a› is phonetic (the donor-oriented one would be أ ‹å›, cf. SA رأس ‹rås›), while س ‹s› is donor-oriented (the self-oriented phonetic graphy would be ص ‹ṣ›).

The oscillation between imitating the SA model and orientation towards the MA phonetic substance reflects the sociolinguistic and pragmatic situation of the spellers, which itself is a sort of attitudinal dualism: They reject SA as a language medium but at the same time accept the Arabic, rather than Latin, script as the graphic medium. As Miller (2017: 91) puts it, "a tension between local authenticity and literary prestige" is the reason why in many cases "writers will opt consciously or not for hybrid systems"[51].

Another spelling principle is responsible for giving up a particular SA spelling which would well reflect MA pronunciation in favour of another graphy. This phenomenon was hinted at by Hoogland (2013a: 73): "it may be the case that cer-

[51] Why should authors writing in MA want to borrow graphic forms from SA rather than use self-oriented spelling is an issue which deserves investigation. In some cases, such as the use of ث ‹ṯ› and ذ ‹ḏ› to mark *t* and *d*, it may be a matter of habit: SA graphies impose themselves on the spellers of MA, among whom there is hardly anybody who would not be primarily a user of written SA. In others, donor form-oriented graphies may be used consciously to facilitate the recognition of a word. For instance, writing the words *hada* 'this' and *dəll* 'he humiliated' in the donor-oriented way, هذا ‹hḏa› and ذل ‹ḏl›, rather than phonetically, هدا ‹hda› and دل ‹dl›, may reflect an intention to distinguish them in writing from *hda* 'he calmed down' and *dəll* 'he indicated', whose phonetic graphy is: هدا ‹hda› and دل ‹dl›, respectively.

tain authors want to stress the non-conformity of *Darija* [MA – M.M.] spelling to MSA [Modern SA – M.M.] orthographic rules". I believe, however, that it is more accurate to interpret such graphies as being aimed at distancing themselves from the SA graphy. Therefore, I will refer to the underlying principle as the *principle of donor-defying*. An example, quoted from Michalski (2017: 238, where spellings resulting from it were referred to as 'anticanonical'), is ا ‹a› marking the final *a* in Form *A* in (2), in contrast to the phonetic graphy ة ‹ħ› in Form *B*, coincident with the donor graphy.

(2)  

‹baġy**a**› (*ṢX* 96[2]) ‹baġy**ħ**› (*ḤB* 18$_{7}$)
baġya 'she wants', lit. 'wishing[f]'
(cf. SA باغية ‹baġyħ› *bāġiya*[tun] 'desiring[f]')

Such a principle should be included in the list of spelling principles of MA, since many graphies cannot be explained otherwise. It is also understandable that writers who decide to give up SA as a medium of literary creation also want to rebel against SA spelling.

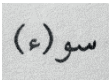
The orthographic principles discussed above are not enough to account for the entire system of MA spelling. A more complete list will be proposed in Ch. V.

Qualitative variation in the MA graphy has been observed rather occasionally in the extant studies, which have been tending to focus on the issue of correspondence between sounds and graphs. As such, research in this area can thus be characterized as having been conducted chiefly within the representational approach (see Sect. III.2.2.1.). While providing us with important insights about how MA is written, it puts us only halfway towards a complete picture of this linguistic phenomenon. By contrast, the primary interest of the present study are the relationships between particular graphs, rather than those between graphs and sounds, with the latter being studied only insofar as this is necessary for the description of the former. With this study concentrating on this issue (signalled in Michalski 2017: 242), it is hoped that a step in another direction in the studies on MA graphy can be made.

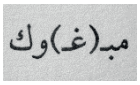
II.2. Internal heterogeneity of Moroccan Arabic

Since MA is not one homogeneous dialect, different pronunciations belonging to different MA dialects may be reflected in texts (cf. Aguadé 2005: 247, 2.2.1.). When analyzing the graphy of a text, it is necessary first to determine which va-

riety of MA it is written in. In most cases, this will be the modern koine or, less frequently, one of the major dialects used in the important Moroccan cities. Determining the variety used should be done on the basis of intratextual rather than extratextual evidence (i.e. the formal features of the text should be considered more conclusive than, for instance, the author's city of origin), because a stylization consisting in imitating other dialects is always possible. Intratextual evidence consists in dialectal peculiarities that a written text evinces in an unambiguous way. The features of *WM*, for instance, indicate that it is written in a northern dialect (Tangier): the allative preposition *n-* 'to' (rather than *l-*, cf. Heath 2002: 238), e.g. طنجة ن ‹n_ṭnǧħ› *n-Ṭanža* 'to Tangier' (*WM* 44); the pronoun *ntina* 'you' (*WM* 18^{4}) (rather than *nta*) or the form منادم ‹mnadm› *mnadəm* 'human being(s)' (*WM* 27_{7}) (rather than *bnadəm*, cf. Heath 2002: 177). In addition to these intratextual arguments, the text is some sort of a sequel to *Al-xubz al-ḥāfī*, a well-known novel written by Muḥammad Šukrī set in some Northern Moroccan cities, including Tangier. In *FL*, the intratextual evidence is the traditional Fez pronunciation, with *ʔ* used instead of *q* and *ġ* instead of *r/ṛ* (cf. Durand 2004: 69), which the speller marked by means of brackets, for instance:

(3) 

‹swo› (*FL* 144^{6})
suʔ 'souq, market'
(cf. *suq* in the modern koine)

(4) 

‹mbġwk› (*FL* 144^{5})
məbġuk 'blessed'
(cf. *məbṛuk* in the modern koine)

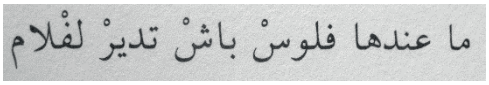
By contrast, the MA passages in *Ḍa.* show features of the Marrakesh dialect, for example, *bit-ək* 'I want you' (*Ḍa.* 47^{10}) is used there rather than *bġit-ək* (cf. Brunot 1931-1952: II, 58; Sánchez 2014: 135) and *hir* 'only' (*Ḍa.* 48_{1}) rather than *ġir* (cf. Sánchez 2014: 240-241). These features alone may not be sufficiently dialect-specific, but, added to the fact that the autobiographic text of *Ḍa.* is set in Marrakesh, they allow us to conclude that these passages are written in this variety. Graphic words which represent such words of limited regional distribution are generally not used as examples in the present study.

II.3. Vocalization in the Moroccan Arabic graphy

Similarly to SA, vocalization signs are not written consistently in MA (Aguadé 2006: 256-257). A text written in this dialect may have no vocalization, but

some are, by contrast, fully vocalized[52]. The majority of MA works, however, is partly vocalized: The vocalization signs are written in some words or with some graphs, the most frequent one of these signs being the *šadda* marking the consonant lengthening. In contrast to SA, MA has no generally acknowledged rules for vocalization. In addition, the three vowel signs borrowed from the SA graphy are insufficient to represent the vocalic system of MA (see Aguadé 2010 for its description). Below, some examples are given.

A fully vocalized text has not been found in the corpus. Only single words represent this feature. What is vocalized seems to be aleatory. The same graphic word may appear in a single text vocalized to different degrees, e.g. *ma bit-š* 'I did not want' (Marrakesh variety) is written in two ways in *Ḍa.*: ما بيتشْ ‹ma_bytš˙› (*Ḍa.* 13[7]) and مَا بِيتْشْ ‹maa_bityt˙š˙› (*Ḍa.* 51[7]). In *BB*, vocalization signs tend to be used only word-finally, for instance:

(5) 

‹ma_ʕndha_flws˙_baš˙_tdyr˙_lf˙lam› (*BB* 108[5])
ma ʕənd-ha flus baš ddir lə-flam
'she has no money to make movies'

Sometimes vocalization is used for semantic disambiguation. For instance, the task of ـُ ‹u› in يضربو العُودْ ‹yḍrbw_alʕuwd˙› *iḍəṛbu l-ʕud* 'they draw lots', lit. 'they hit the wood' (*ḤB* 25$_8$) is to prevent a less expert reader from reading the unvocalized graphic word العود as *l-ʕəwd* 'the horse', even though the reading 'they hit the horse' would make little sense in this context. Rarely, vocalization is used to distinguish MA from SA, for example, the *kasra* in كِما كان ‹kima_kan› (*BB* 70[10]) is the only thing to indicate that the graphic phrase represents MA *ki-ma kan* 'as he/it[m] was', not SA *ka-mā kāna* 'idem'. However, more often that not, vocalization is redundant. For example, only one reading would be possible for وتْزِيزْنْتْ ‹wt˙zi yz˙n˙t˙› *u-ḍziẓənt* 'and I became speechless' (*ḤB* 39$_6$) in its context even without vocalization signs.

Another conspicuous feature of vocalization used in the sources examined is that it is often applied in a way contrary to expectations a reader may have developed on the basis of its use in SA. An example of a vocalization sign that appears to serve goals other than those in SA is ـْ ‹˙›, the *sukūn*, written quite frequently where *ə* or *ŭ* are pronounced (cf. Aguadé 2006: 256), e.g. بْعْدَه ‹b˙ʕ˙da h›

[52] Aguadé (2006: 256) considers this "absurd choice that helps only to make reading of such texts more difficult" to be "a Moroccan peculiarity", which he contrasts with EA, where vocalization is "very unusual" (cf. Rosenbaum 2004: 294).

bəʕda ‘already; first’ (*Mn.* 44$_6$), تَياخُذ ‹tᵃyax·ḏ› *ta-yaxŭd* ‘he takes’ (*Ḍa.* 47^7), كيدوُّر ‹kydw·r› *ka-yḍŭwwəṛ* ‘he searches’ (*Fʕ* 62^2). Such graphies probably reflect the phonological awareness of the MA spellers, according to which there is no vowel there (cf. Aguadé’s opinion in Michalski 2016: 387, footnote 4), although scholars (Benhallam 1980, 1990, as cited in Boudlal 2011) have posited a *shwa* splitting initial three-consonant clusters. An alternative explanation might be based on Durand’s view (2004: 61-64) according to which the MA phonology does allow words to begin with a three-consonant cluster, hence, at least in some cases, the *sukūn* might mark the real absence of a vowel. A similar issue is the use of a *sukūn* after a geminate, where, according to the SA usage, one could expect a *sukūn* with a *šadda*, e.g. يدُوْزوا ‹ydᵘw·zwa› *ydŭwwzu* ‘they spend’ (*BB* 107^5), تَتْحَبْ ‹tᵃt·ḥᵃb·› *ta-tḥəbb* ‘she loves’ (*Ḍa.* 41$_1$), as if gemination were not perceived as such by the spellers of MA (cf. Durand 2004: 75)[53]. Spellings such as these are quite often and cannot be considered mere misprints. They rather show new functions of vocalization signs developed within MA.

Although most vocalization signs are redundant, some of them, those denoting vowels that otherwise are signalled by primary graphs which function as *matres lectionis*, can be said to be functionally equivalent to primary graphs, e.g. رَجل ‹rᵃğl› vs. راجل ‹rağl› *ṛažəl* ‘man’[54]. However, the use of vocalization signs in this function is restricted: primary graphs can be used in any position of a graphic word while vocalization signs normally mark vowels word-medially (since they always follow a graph, they cannot be used word-initially). Vowels signalled by a vocalization sign word-finally are rare; the most frequent of them seems to be ‹ⁱ› marking *-i* in feminine pronouns, e.g. أنتِ ‹åntⁱ› *nti* ‘youᶠ’ (*BB* 44^6) and in 2nd person singular past verbs, e.g. قلتِ ‹qltⁱ› *qŭlti* ‘you said’ (N 60$_{11}$), as opposed to أنتي ‹ånty› (*ʕD* 32$_7$) and قلتي ‹qlty› (*Ṣa.* 215^4) written with a final ي ‹y›. There are only isolated cases of the word-final use of ـُ ‹ᵘ› or ـَ ‹ᵃ›, e.g. in عمرُ ما ‹ʕmrᵘ_ma› *ʕəmmṛ-u ma* ‘he never’ (*TN* 8$_6$) and شوِيَ ‹šwⁱyᵃ› *šwĭyya* ‘a little’ (*TM* 54^5), as opposed to the usual و ‹w› and ة ‹ħ›, respectively, as in عمرو ما ‹ʕmrw_ma› (*TM* 49$_2$) and شوية ‹šwyħ› (*Aʕ* 42^7). The *šadda*, too, can be functionally equivalent to a primary graph because it can be written word-medially instead of a repeated consonant graph to signal the lengthening of a consonant or semivowel. For example, the geminated *l* in *məlli* ‘when’ is marked by means of a *šadda* in مِلّي ‹mⁱl·ⁱy› (*Mm.* 116^4), and as ‹ll› in مللي ‹mlly› (*Ṛḥ.* 86$_4$). Sometimes, these co-functional units co-occur, for instance in

[53] In contrast to the absence of an expected *šadda*, this sign is sometimes written with no justification when a consonant is not lengthened, e.g.: دابّة ‹dabˑħ› *daba* ‘now’ (*BB* 44$_7$), نَّكْلُو ‹nˑᵃkˑlᵘw› *naklu* ‘we eat’ (*TM* 66^7).

[54] Forms without a vocalization sign or a *mater lectionis* are also currently used, e.g. رجل ‹rğl› ‘idem’.

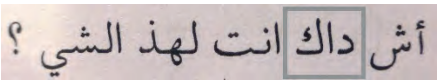
ملّي ‹mll·y› (*BB* 44_{10})[55]. Such redundancy may be caused by the superposition of two spelling principles if gemination is a result of assimilation. For instance, in the graphic word نضّت ‹nḍ·t› (*Ṛḥ.* 148_{11}) for *nəḍḍt* 'I got up', pronounced [nəṭṭ], the *šadda* signals gemination phonetically, while the graphy ‹ḍt› is morphological[56]. Another case of the *šadda* having a function similar to that of a primary graph is its being written after ب ‹b› to indicate that the sound *p* is intended, e.g. لابّيسين ‹lab·ysyn› *lapisin* 'swimming pool' (*Ṛḥ.* 175_{8}). As already mentioned before, although some of the vocalization signs are co-functional with primary graphs, they are not taken into consideration in the present work.

II.4. The context-dependence of Moroccan Arabic graphy

Like SA graphy, the graphy of MA is context-dependent, i.e. a context is needed for some graphic words to determine their intended meaning and, consequently, pronunciation. Two principal reasons for this situation are specifically related to Arabic script. The first of them is that some sounds are not represented graphically. The second one is the ambiguity of some graphs and the homography of some sounds, caused chiefly by the absence of special graphs for marking some Moroccan sounds in Arabic script. A linguistically universal factor is the existence of homophones, which in most cases are also homographic. In (6)-(8), examples of homographs are given with two meanings and two pronunciations which cannot be determined without resorting to context.

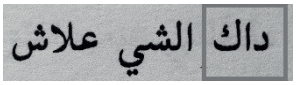
(6) داك ‹dak›:

dda-k 'it took you'



‹åš_**dak**_ant_lhḍ_alšy› (*TM* 67_{2})
*aš **dda-k** nta l-had š-ši?*
'What business is it of yours?',
lit. 'What took you to this?'

dak 'that'



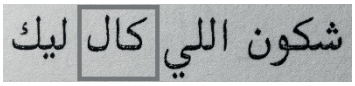
‹**dak**_alšy_ʕlaš› (*QQ* 47_{10})
***dak** š-ši ʕlaš…*
'that is why…', lit. 'that thing why'

[55] Forms with neither *šadda* or a repeated graph, such as ملي ‹mly› (*TM* 42^{2}), as well as forms with neither *šadda* or ال ‹al› marking the definite article morphologically, such as هاذاك شي ‹haḍak_šy› *hadak š-ši* 'that', lit. 'that thing' (*Ḍ* 100_{12}), can be found too.

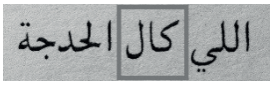
[56] This practice was observed also in SA, by Badawi *et al.* (2004: 15).

(7) **كال** ‹kal›:

gal ‘he told, he said’ *kal* ‘he ate’



‹škwn_ally_**kal**_lyk› (*ʕD* 202$_{10}$)
*škun lli **gal** li-k*
‘Who told you’?

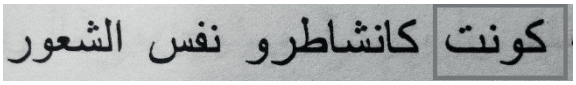


‹ally_**kal**_alḥdğħ› (*ḌK* 8$_{9}$)
*lli **kal** l-ḥədža*
‘The one who ate a colocynth[57]’

(8) **كونت** ‹kwnt›:

kŭwwənt ‘I hid and waited’ *kŭnt* ‘I was’

‹wåna_**kwnt**_ʕlyh› (*TN* 18^{12})
*w-ana **kŭwwənt** ʕli-h*
‘and I hid and waited for him’

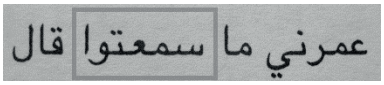


‹**kwnt**_kanšaṭrw_nfs_alšʕwr› (*Ṛḥ.* 24$_{1}$)
***kŭnt** ka-nšaṭṛ-u nəfs š-šʕuṛ*
‘I reciprocated his feelings’

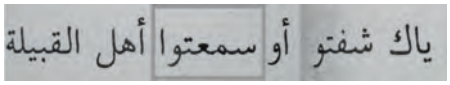
(9)-(11) show examples of homographs with two different meanings but the same pronunciation.

(9) **سمعتوا** ‹smʕtwa›:

səmʕt-u ‘I heard him’ *səmʕtu* ‘you^{pl} heard’



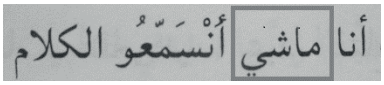
‹ʕmrny_ma_**smʕtwa**_qal› (*TN* 18$_{3}$)
*ʕəmməṛ-ni ma **səmʕt-u** qal…*
‘I never heard him say…’



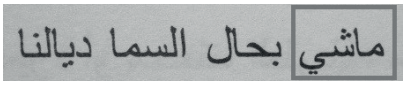
‹yak_šftw_åw_**smʕtwa**_åhl_alqbylħ› (*Ṣa.* 230$^{1-2}$)
*yak šəftu aw **səmʕtu** ʔahl lə-qbila*
‘Have you^{pl} seen or heard the people of the tribe?’

[57] The bitter fruit of colocynth is a symbol of sorrow in Arabic culture.

(10) **ماشي** ‹mašy›:
maši ‘going to’ (future marker) *ma-ši* ‘is not’

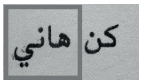


‹åna_**mašy**_åᵃnˑsᵃmːʕᵘw_alklam› (*MX* 16[7])
*ana **maši** nsəmmʕ-u l-klam*
‘I will give him a piece of my mind’

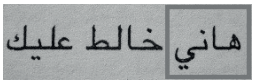


‹**mašy**_bḥal_alsma_dyalna› (*Ṛḥ.* 163[5])
***ma-ši** bḥal s-sma dyalna*
‘it is not like our sky’

(11) **هاني** ‹hany›:
hani ‘peaceful, calm’ *ha-ni* ‘presentative + I’



‹kn_**hany**› (*QQ* 63[4])
*kun **hani***
‘Don’t worry!’, lit. ‘Be calm!’



‹**hany**_xalṭ_ʕlyk› (*ṬN* 30[1]).
***ha-ni** xaləṭ ʕli-k*
‘I’ll be right with you’

Generally, homography in MA is not a problem as far as determining the meaning is concerned, if the context is given. Determining the pronunciation intended by the speller may be however impossible in some cases (see Sect. IV.2.1.).

III. Methodology and theoretical considerations

In this chapter, graphetic, graphemic and other linguistic concepts used in the subsequent parts of this work will be introduced and illustrated with examples taken from MA. Graphetics is understood here as a branch of linguistics dealing with graphs[58], i.e. units of written language that are distinguishable on a purely formal basis without taking into consideration relations based on their meaning-differentiating functions[59]. The latter relations are the object of graphemics (cf. Günther 1988: 64, 71). This distinction between graphetics and graphemics is, at least at some points, analogous to that between phonetics and phonemics[60].

First, graphetic concepts such as the actual graph, homography (in the broad sense), graph, heterography and continuous graphic words are introduced. This is followed by a presentation of graphemic concepts such as discontinuous graphic words, distinctive and variational heterography, with its three types: qualitative, quantitative and linear, as well as graphemic variation, graphemic opposition and grapheme.

---

[58] Recall that the term 'letter' is reserved for any of the 28 elements of the Arabic alphabet (abjad). In adopting the term 'graph', I follow Kohrt (1986: 90), who, arguing for a graphematic terminology that is as close as possible to the phonemic one, says: "since the term 'phone' is normally understood to designate a particular instance of a 'phoneme', we should stick to a parallel interpretation of the term 'graph', i.e. a graph has to be considered as a materialization of a certain grapheme in a concrete written utterance". In some contexts, however, the terms 'letter' and 'graph' can be used as synonyms in the present study, without this leading to misunderstandings.

[59] Different conceptual and terminological decisions can obviously be made in this respect. For instance, in the model proposed by Hammarström (1966: 58), as long as pronunciation is not taken into consideration, the units of analysis are referred to as 'type' and 'typeme', while the concepts of 'graph' and 'grapheme' are used when the text description is connected with available knowledge about the pronunciation. Hammarström understands 'types' as "variants of the letters", which are "the smallest written segments capable of differentiating significations" (1981 [1964]: 91), and 'graph' as "a typeme or a combination of typemes which denote a phoneme" (1981 [1964]: 96).

[60] The term 'graphemics' is chosen here rather than 'graphology', used, for instance, by Crystal (2008: 220). Although the latter, ending in '-logy', has the virtue of indicating that it is to graphetics what phonology is to phonetics, it is too strongly associated with the popular concept of the analysis of handwriting.

III.1. Graphetic concepts

When looking at a text written in Arabic script, or in any other alphabet or abjad, what we are faced with are actual graphs, i.e. those that can be seen on a particular medium: usually printed on a particular piece of paper or displayed on a particular screen[61]. Actual graphs that differ only with respect to such features as size, colour, the kind of printing type (italic, bold), font or style, i.e. features which are deemed irrelevant by users for the identification of the (abstract) graphs, are said to be bound by the relation of *homography* (in the broad sense, i.e. irrespective of the meaning and pronunciation). For instance, all actual graphs shown in Table 2, representing various printing types and fonts, are homographic and belong to a single (abstract) graph غ̇ ‹ġ›[62] (the fact that it is used in SA graphy to represent, in most cases, the sound *ġ* is irrelevant at this point).

Table 2. *Various actual graphs belonging to graph* غ̇ *‹ġ›*

غ̇	غ̇	غ̇	غ̇	غ̇
Times New Roman, roman light	Times New Roman, bold	Times New Roman, italic	Andalus	Arial Unicode MS

In practice, the distinction between actual graphs and (abstract) graphs is not needed in most cases and using the term ‘graph’ will not lead to misunderstandings in our further considerations.

Actual graphs or graphs that are not homographic are *heterographic*. This is, for instance, the case of the pair غ̇ ‹ġ› and ق ‹q›. The relation of heterography is discussed in more detail in subsequent paragraphs.

In order to be able to analyze the relations between graphs, i.e. basic units of the present description, we need to clarify what is and what is not considered a graph. The SA alphabet composed of 28 letters can be used as a point of departure but is not an exhaustive list. The set of graphs must be understood in a broader sense and will also include extra-alphabetic graphs. These are, on the one hand, graphs used in CA and SA but not counted as separate letters: the *hamza* sign with no support, ء ‹o›, and its combinations with other letters functioning as its support:

[61] The idea of distinguishing graphs from actual graphs is based on that of the distinction between actual phones and phones proposed in Bańczerowski *et al.* (1982: 121-123).

[62] They would be considered allographs by some authors, e.g. Rogers (2005: 11) and Crystal (2008: 20).

أ ‹å›, إ ‹ą›, ؤ ‹ẘ›, ئ ‹ŷ›, including آ ‹ã› (the *ʔalif madda*)[63], as well as ى ‹á› (the *ʔalif maqṣūra*, i.e. the dotless *yāʔ*) and ة ‹ħ› (the *tāʔ marbūṭa*, i.e. the dotted *hāʔ*). On the other hand, the extra-alphabetic graphs comprise six graphs not used in the standard classical version of SA spelling but encountered in texts written in Arabic dialects, less frequently in various local varieties of SA, to denote sounds absent from CA and SA pronunciation (they are also used for writing other languages). These graphs are: پ ‹p›, چ ‹č›, گ ‹ǵ›, ڴ ‹ĝ›, ڤ ‹v_f› and ڤ ‹v_q›. Finally, there are two graphs belonging to the local Moroccan variety of Arabic script, i.e. the Maghrebi script: ڢ ‹f› and ڧ ‹f_q› (the latter being the one of the two shapes of ف ‹f› which is used in the non-neutralized position).

Ligatures (see Sect. I.2.), which are phenomena pertaining only to the graphetic or typographical level, are considered irrelevant in this work. In transliteration, they are solved into their constituting graphs. For instance, the only obligatory ligature لا is transliterated as ‹la›.

Being entirely predictable and of little theoretical importance, the differences of the shapes of graphs depending on the graphic environment, i.e. whether a given graph is isolated, connected to the right and/or left, are disregarded. However, some comments are necessary about the issue of the *positional neutralization of heterography*, of which three instances have been identified in the course of this work. This term means that a pair of actual graphs describable as belonging to two different graphs on account of their formal differences in a *non-neutralized* position in a graphic word shows no formal difference in a *neutralized* position – and should thus be considered belonging to a single graph. More specifically, in the three pairs identified, the non-neutralized position is 'non-connected to the left' and the neutralized position is 'connected to the left'. In the discussion which follows, these actual graphs are represented in their shapes used in the non-neutralized positions.

Two of these three pairs share a relevant feature which enables them to be discussed together: Their *rasm*s (bare shapes without dots) in the non-neutralized position are ڧ and ق, i.e. have a shallow versus deep bowl (a description borrowed from Daniels 2013: 413). Generally, the shallow-bowled *rasm* ڧ is used as the basis for:

---

[63] Following this approach, the distinction between such units as, for instance, ا ‹a› and أ ‹å›, is qualitative. In an alternative description, one could assume that each of the following graphs, ا ‹a›, و ‹w› and ي ‹y›, can be combined with the graph ء ‹o›, to result in أ ‹å›, إ ‹ą›, ؤ ‹ẘ› and ئ ‹ŷ› (in the case of ي , this combination would automatically require the deletion of the two dots). Thus, for instance, أ ‹å› would be described as composed of two graphs, the *ʔalif* ا ‹a› functioning as the supporting graph and the *hamza* sign ء ‹o›. Consequently, the opposition between ا ‹a› and أ ‹å› would be quantitative. This approach has been rejected as descriptively more troublesome and less adequate.

a. ف – marking *f* (and *v*, in borrowings) in general Arabic graphy and
b. ڢ – marking *f* in local Maghrebi graphy.

The deep-bowled *rasm* ى is the basis for:

a. ق – marking *q* (and *g*, in borrowings) in general Arabic graphy and
b. ى – marking *q* in local Maghrebi graphy.

Since the graphs based on the shallow-bowled *rasm* ف mark fricatives (*f* and *v*), while those based on the deep-bowled *rasm* ى mark plosives (*q* and *g*), one could speak of the 'fricative *rasm*' and 'plosive *rasm*'. It might be concluded that the shallow-bowled graph ڤ marks *v* (fricative) while the deep-bowled ڧ marks *g* (plosive). However, the writing practice revealed in the corpus does not confirm such correspondence (see under "ق ‹q› Rv ڤ ‹v›" and "GRAPH ڤ ‹v›" in Ch. VI).

Having unique pointings, the graphs ڢ , marking *f* in local Maghrebi graphy, and ق , marking *q* in general Arabic graphy, are never affected by the positional neutralization of heterography. The two remaining pairs, ف vs. ى and ڤ vs. ڧ , need to be discussed. Within the first of these pairs, the graph ف ‹f_f› (the Ḥafṣ tradition) marks *f* in general Arabic graphy, whereas ى ‹f_q› (the Warš tradition) marks *q* in the local Maghrebi graphy (see Sect. I.1.). The neutralization of heterography between them is exemplified in فال , a graphic word which represents MA *fal* 'omen' if it is read according to the Ḥafṣ tradition and *qal* 'he said' if the Warš tradition is followed. Such ambiguities do not happen if the two traditions are combined in a sensible way[64]. However, a problem arises in a macrodescription including the totality of graphs used in both traditions.

In consideration of our initial assumption that graphs are identified exclusively by virtue of their formal features, ف and ى cannot be distinguished as two different graphs in the neutralized (connecting) position, where each of them has the shape ـفـ. For this reason, a single graph ـفـ is posited in the description, transliterated ‹f›[65], with two different non-linking shapes: ف and ى. The latter are differentiated from each other in transliteration by means of subscripts reflecting the (typical)

[64] The combination of the Warš graphy of ڢ for *f* with the Ḥafṣ graphy of ق ‹q› for *q* (this is the case of the SA and MA text in *Ḥaš.*) does not result in a dysfunction. By contrast, combining the Ḥafṣ graphy ـفـ ‹f› and ف ‹f_f› for *f* with the Warš graphy ـفـ ‹f› and ى ‹f_q› for *q* would be dysfunctional because of the neutralization of heterography between these two graphs in the connecting positions.

[65] It seems impossible to propose a transliteration symbol which would reflect both phonetic functions of this graph. Ultimately, I have decided that ‹f›, rather than ‹q›, is a more intuitive choice, because marking *f* is by far more frequent a function of ـفـ than marking *q*, the latter being marginal, as the corpus reveals.

phonetic value of their unconnected *rasm*s: ‹f_f› and ‹f_q›, respectively[66]. The same solution is adopted for ڤ and ڧ , a pair which is a differently dotted analogon of ف and ڧ. They too are formally differentiable only in the non-connecting position and are thus treated as one graph, ڤ ‹v›[67], with two non-linking shapes: ڤ and ڧ. Their differentiation from each other in transliteration is achieved by means of subscripts: ‹v_f› and ‹v_q›, respectively.

The third case of the positional neutralization of heterography is a graph which in the neutralized position has the form ڪ , i.e. that of a *kāf* with three dots above. Its function is to mark *g*. In the non-connecting position, two forms can be taken into consideration: ک and ڪ. They are based on the two non-connecting *rasm*s of the *kāf*: ک and ك, described by classical Arab grammarians in the following way: The first of them (used nowadays in modern Persian graphy) is termed *al-kāf al-maškūla*, i.e. one 'provided with a straight line [*šakl* – M.M.] above its stem' (Gacek 2001: 79) serving as a diacritic mark which distinguishes it from word-final ل ‹l› (cf. al-Qalqašandī [n.d]: 80-81, 155). Ibn Durustawayh, a classical philologist (d. 956), disapproves of using this final *rasm* (1921: 72-73). The second *rasm*, ك, used currently in SA graphy, has no such line (is *muʕarran min al-šakl* 'lacking the *šakl*', Ibn Durustawayh 1921: 67). Instead, it has a little Arabic ک inside it to distinguish it from ل ‹l› in the word-final position. Which of these two non-neutralized *rasm*s is the base for the *kāf* with three dots in MA is not obvious. According to Kew (2005: 4-5), "the base form used is consistently ک and not ك". He observes that "there are sources that show authors deliberately using a final *kashida*[68] to force a ك-shaped character to take on the shape ک, so strong is the feeling that ک rather than ڪ is the correct form (…)". Unfortunately, we do not know what sources Kew meant. The form indicated in Aguadé (2006: 259) is ڪ , while in Caubet (2017: 134), by contrast, it is ک , or, to be precise, ګ, with a different shape of the bottom stroke (probably a result of the use of a different font). As for the usage in the corpus, the graph in question appears only in *ḌK* and *ḤB*. However, in the analyzed fragments of the

---

[66] It could be argued that the actual graphs ف and ڧ , which I decided to consider unconnected forms of a single graph ڤ ‹f›, should be described as two different graphs because they are formally distinguishable. This is true. However, such a solution would make the description too complex, with as many as three graphs having to be distinguished: ڧ , ف (each of them occurring only in non-neutralized positions) and ڤ (occurring only in neutralized positions).

[67] A transliteration symbol reflecting both phonetic functions of this graph would have to be some sort of a combination of ‹v› and ‹q›. I have opted for ‹v› because marking *v* rather than *q* is its more frequent function.

[68] The term *kashida* (Arabic كشيدة *kašīda*, borrowed from Persian كشيده *käšidä* 'extended', 'elongated') is used in typography and computing to denote the character ‹ـ›, marking a horizontal elongation of a letter in a word printed with the use of Arabic script (in this sense, it is also called *taṭwīl* 'elongation'). Originally, in Arabic calligraphy, it means such an elongation in a handwritten word.

former source, it does not occur in a non-neutralized form. In the latter, wherever a non-neutralized form should be used, its neutralized connecting form ڴ appears, e.g.:

(1) 

‹fwĝ› (*ḤB* 16[8])
fug ‘over, upon’

This probably confirms Kew’s opinion (although the solution adopted in *ḤB* was to use a Unicode connecting presentation form rather than to add a *kashida*). Since more corpus data are needed to decide on the relationship between ڪ and ڪٛ, I transliterate ڴ as ‹ĝ› without any further distinction.

Following the above considerations, a list of 42 graphs has been established, including the 28 letters of the Arabic alphabet and 14 graphs not contained therein, listed in Table 1 in Ch. ‘Transliteration’.

A *space* is understood as a spatial distance between two graphs that is perceived as such by the users of MA graphy. Whether a space is there or not is easy to determine in a position following a connecting graph because this is visible in the shape of the graphs: For instance, ا م ‹m_a› is written with a space, while ما ‹ma› is not. By contrast, in the case of non-connecting graphs the issue is a matter of only a relative distance, cf. ام vs. ا م, which may be dubious. Another problem is that connecting graphs can sometimes be found written very close to the following graph but in their non-connecting form, e.g. ما. However, since in this study only qualitative relations are dealt with, no special distinctions need to be introduced for such phenomena for the time being.

A generic term covering both graphs and the space used in this description is *characters*.

Graphic units that are not bound by homography are bound by *heterography* (symbolically, *Hg*)[69] and are *heterographs* with respect to each other. This relation may bind both graphs, e.g. غ ‹ġ› with ع ‹ʕ›, and graphic words (examples below). It is defined on purely formal criteria and obtains regardless of the relationhips of meaning, pronunciation and morphology. This is visible in the translations and

[69] This term is not to be understood in the conventional manner, as defined, for instance, by the *Oxford English Dictionary*, as “Spelling that differs from that which is correct according to current usage; ‘incorrect’ spelling”. Its sense is closer to, but not identical with, the second meaning indicated there: “Irregular and inconsistent spelling (as the current spelling of English)” (“Heterography, *n.*” 2018).

transliterations given, not always exhaustively, for the graphic words in the following examples:

(2)	من كم ‹mn_km› *mən kŭmm* ‘from a sleeve’	***Hg***	منكم ‹mnkm› *mən-kŭm* ‘from you[pl]’		
(3)	ما كانش ‹ma_kanš› *ma kan-š* ‘he was not’	***Hg***	ماكانش ‹makanš› *ma kan-š* ‘he was not’		
(4)	ديل ‹dyl› *dil* ‘tail’	***Hg***	ديال ‹dyal› *dyal* ‘of’ (possessive preposition)	***Hg***	اديال ‹adyal› *dyal* ‘of’ (possessive preposition)
(5)	بيت ‹byt› *bit* ‘room’	***Hg***	بير ‹byr› *bir* ‘well’	***Hg***	بئر ‹bŷr› *bir* ‘well’

Three types of heterography can be distinguished depending on the type of the distinctive unit which differentiates two graphic words. In (2)-(3), the distinctive unit is the space as opposed to its absence. This type of heterography is termed *linear heterography*. In (4), the distinctive unit is a graph as opposed to its absence. This type of heterography is referred to as *quantitative*[70]. In (5), the distinctive unit is graph *x* in a given position of the graphic word as opposed to graph *y* in the same position. This type is termed *qualitative heterography*.

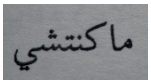
The *pronunciation* of a graphic unit (a graph, a graphic word etc) is a real or potential phonetic form that can be associated with it[71]. For instance, the

[70] Linear heterography and quantitative heterography can be both viewed as subtypes of quantitative heterography in the broad sense because each of them relies on the opposition ‘presence of a character’ vs. ‘absence of a character’, the respective distinctive character being a space or a graph.

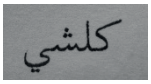
[71] The phonetic form is understood here as systemic, not impressionistic. As has been mentioned at an earlier point, broad transcription is used throughout this work as long as the phonetic transcription is not necessary.

graphic word كتبت ‹ktbt› ‘she wrote’, the meaning of which is assumed here to be known from context, has two pronunciations: *kətbət* and *kətbat*, with the vowel *a* in *kətbat* being not represented in writing. By contrast, the graphic word كتبات ‹ktbat› ‘idem’ has only one pronunciation: *kətbat*, because the vowel *a* is explicitly signalled in writing. Another example is the graphic word قال ‹qal› ‘he said’. It has two pronunciations: *qal* and *gal*, used interchangeably in some regions of Morocco. By contrast, گال ‹ǵal› ‘idem’ has only one pronunciation, *gal*, because the graph گ ‹ǵ› is used specifically to represent *g*. A graphic unit (a graph, a graphic word, etc.) which has more than one pronunciation is called *phonetically ambiguous*, while those with only one pronunciation are *phonetically univocal*[72]. Phonetic ambiguity denotes the mere fact of having more than one pronunciation, irrespective of whether the pronunciation intended by the speller can or cannot be identified by the reader, e.g. with the help of the context (as opposed to phonetic indeterminacy, see Sect. IV.2.1.).

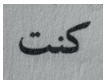
Graphic words[73] can be continuous or discontinuous. A *continuous graphic word* is a graphic unit composed of graphs without spaces (punctuation marks are disregarded). Examples:

(6) 

‹makntšy› (*MḤ* 45[13])
ma kŭnt-ši
‘I was not’

(7) 

‹klšy› (*DB* 251[1])
kŭll ši
‘everything’

(8) 

‹knt› (*Mu.* 47[13])
kŭnt
‘I was’

Discontinuous graphic words are defined on the basis of semantic and phonetic properties of continuous graphic words. Hence, they are discussed in the following chapter, devoted to graphemic concepts.

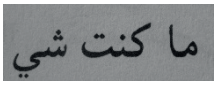
[72] A need to introduce the concept of the set of all pronunciations which can be associated with a graphic unit was perceived by Hammarström (1976: 115), who proposed the term ‘nunceme’ to denote “[a]ll pronunciations of a letter (typeme), or of certain combinations of letters (…)”.

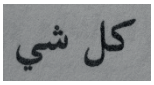
[73] A distinction between (abstract) ‘graphic words’ and ‘actual graphic words’ can be made, on analogy to that between (abstract) ‘graphs’ and ‘actual graphs’. In practice, however, such a distinction would seldom be needed.

III.2. Graphemic concepts

With the characters, viz. graphs and spaces, being analyzed with respect to their meaning-differentiating and pronunciation-differentiating functions, the functional aspects of MA graphy come into play. This is thus the point where the graphetics meets semantics and phonetics, and where the analysis is shifting from the graphic level to the graphemic level[74].

The first concept to be introduced here on the basis of the semantic and phonetic properties is that of a *discontinuous graphic word*. It is understood as a linear sequence of two – or more – continuous graphic words divided by a space – or spaces – and for which a homosemantic (having the same meaning) and homophonous (having the same pronunciation) continuous graphic word (divisible continuous graphic word) exists[75]. Examples:

(9) 
‹ma_knt_šy› (*MX* 18_{11})
ma kŭnt-ši
‘I was not’

(10) 
‹kl_šy› (*ḤB* 35^{3})
kŭll ši
‘everything’

The discontinuous graphic words in (9) and (10) have their homosemantic and homophonous correspondents in discontinuous graphic words shown in (6) and (7) in the preceding section. For the continuous graphic word given there in (8), no corresponding dicontinuous graphic word exists.

On the basis of the semantic and phonetic relations between graphic words, heterography can be divided into *distinctive* and *variational*. *Distinctive heterography* (symbolically, *DH*) binds heterographic words that differ, even context-independently, in meaning and pronunciation. The examples (2′), (4′) and (5′) given below, based on (2), (4) and (5) shown above, illustrate this relation in its three types: linear distinctive heterography (2′), quantitative distinctive heterography (4′) and qualitative distinctive heterography (5′).

---

[74] Note that in the present approach to MA graphy, graphemics is understood in its broad sense, i.e. as dealing not only with the distinctive functions of graphs as opposed to one another (qualitative graphemics) but also graphs as opposed to their absence (quantitative graphemics) and the space as opposed to its absence (linear graphemics).

[75] Here and in other parts of the present work, what is understood by such concepts as “exists”, “can be substituted for”, “can be added”, “can be deleted”, etc., is that a given form has been recorded in the corpus or, in a looser sense, can be predicted as an expected form on the basis of analogy; cf. the concept of expected forms introduced at the end of this chapter.

(2′)	من كم	***DH***	منكم
	‹mn_km›		‹mnkm›
	mən kŭmm		*mən-kŭm*
	‘from a sleeve’		‘from you^{pl}’
(4′)	ديل	***DH***	ديال
	‹dyl›		‹dyal›
	dil		*dyal*
	‘tail’		‘of’ (possessive preposition)
(5′)	بيت	***DH***	بير
	‹byt›		‹byr›
	bit		*bir*
	‘a room’		‘a well’

Variational heterography (symbolically, *VH*) binds heterographic words which are homosemantic and homophonous[76] (even though for some graphic words, this can be determined only with the help of context[77]). Such words are called *variational heterographs* (or simply *heterographs*, if this shortening does not involve any ambiguity). They are the principal object of our interest in this work. Below, some graphic words used in (3)-(5) are given to illustrate its three types: linear variational heterography (3″), quantitative variational heterography (4″) and qualitative variational heterography (5″). Each of these types is discussed in more detail in Sect. III.2.1.

(3″)	ما كانش	***VH***	ماكانش
	‹ma_kanš›		‹makanš›
	ma kan-š ‘he was not’		
(4″)	ديال	***VH***	اديال *dyal*
	‹dyal›		‹adyal›
	‘of’ (possessive preposition)		

[76] Heterographs which are homophonous but not homosemantic also exist in MA but they do not concern us here.

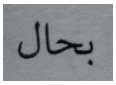
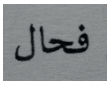
A third condition for graphic words to be considered bound by variational heterography could be added: that of having the same morphological structure. For SA, it would exclude such pairs of graphic words as verbs نما ‹nma› and نمى ‹nmá›. They both have the same meaning, ‘he/it^{m} grew’, the same pronunciation, but different morphological structures, which becomes visible in other inflectional forms. This is, however, a rare phenomenon and can be disregarded for MA.

[77] Cf. the discussion of neutralized graphemic relation between graphs in Sect. III.2.2.2.

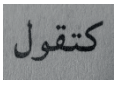
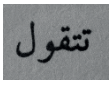
(5″) بير ***VH*** بئر
‹byr› ‹by̆r›
bir ‘a well’

Some comments on the conditions of the variational heterographs being homophonous and homosemantic are needed. As for homophony, the pronunciations of the heterographs do not have to be absolutely identical as long as the phonetic differences do not influence the graphy in question (this concerns, for instance, affricatization, not marked in writing). In some cases, the condition of being homophonous does not mean that the two graphic words must have *exclusively* identical pronunciations. They only need to have *at least* one shared identical pronunciation. For instance, in the pair گال ‹ǵal› and قال ‹qal›, the pronunciation of the former is *gal*, while the latter can be pronounced either *gal* or *qal*. If no extra- or intratextual circumstances indicate which of these two pronunciations was intended by the speller, the two graphic words are considered homophonous – because both are pronounceable as *gal*.

The condition of homophony excludes from the relation of variational heterography the following pairs of graphic words:

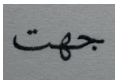
(11)  
‹**b**ḥal› (*Ṣa.* 209[12]) ‹**f**ḥal› (*WM* 25$_6$)
b*ḥal* ***f****ḥal*
‘like, similar to’ (preposition)

Both these graphic words mean the same and, historically, are composed of the noun *ḥal* ‘state, situation’ with a preposition meaning ‘in’. However, the prepositions are different, *b-* and *f-*, which results in different pronunciations, depending on the dialect variety (cf. Heath 2002: 235-36).

(12)  
‹**k**tqwl› (*MX* 131[1]) ‹**t**tqwl› (*LŠ* 53$_2$)
ka*-dgul* ***ta****-dgul*
‘you say’

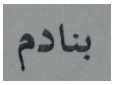
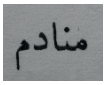
Both graphic words have the same meaning and the same morphological structure: a present verb with an aspectual preverb. The preverb, however, has two

different phonetic forms: *ka-* and *ta-*, the use of which depends on geographical distribution (cf. Kjamilev 1968: 44, footnote 80; Aguadé 1996: 198-199)[78].

(13)  

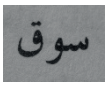
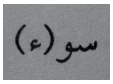
‹ğh**ħ**› (*Ḍḥ.* 174_{10}) ‹ğh**t**› (*Mn.* 23_{3})
*žih**a*** *žih**t***
‘direction, place, side’

These graphic words have the same meaning but different pronunciations due to their different syntactic status: Form *A* is used in the free state (e.g. in the expression *kŭll žiha* ‘everywhere’) and its feminine suffix is *-a*. By contrast, *B* is used in the construct state (e.g. in the expression *mən žiht* ‘on the part of’) and its feminine suffix is *-t*.

(14)  

‹**b**nadm› (*ḤB* 9^{11}) ‹**m**nadm› (*WM* 27_{7})
***b**nadəm* ***m**nadəm*
‘human being(s)’

These two graphic words both mean the same and are historically related with SA *ʔibn ʔādam* ‘human being’, lit. ‘son of Adam’. The form *bnadəm* is the dominating form in Morocco, while *mnadəm*, resulting from the nasalization of *b* caused by the neighbouring *n*, is used only in some varieties (cf. Heath 2002: 177).

(15)  

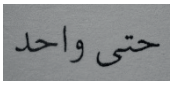
‹sw**q**› (*ḤB* 44^{13}) ‹sw**o**› (*FL* 144^{6})
*su**q*** *su**ʔ***
‘souq, market’

[78] According to Youssi (1992: 61), the two different preverbs have ceased to mark any regional or dialectal differentiation and have become free variants, although *ka-* may seem preferrable to some users and, in addition: “dans une suite de verbes d’un même énoncé (coordonnés ou subordonnés), le premier a la variante *ka* et un ou plusieurs des verbes suivants peuvent avoir *ta*”.

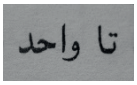
These two graphic words have the same meaning. Form *A* has the pronunciation *suq*, which is dominant in Morocco, while *B* is pronounced *su?*, i.e. with a glottal stop instead of *q*, a trait characteristic of many Jewish and some Muslim urban pre-Hilalian dialects (Aguadé 2003: 87), like that of Fes, the pronunciation of which is represented in *FL* (note the brackets added by the speller to mark this local peculiarity).

It is more difficult to decide how to treat graphic words which represent pronunciations resulting from phonetic reductions, especially consonant-cluster reductions, whether occurring in rapid speech or having occurred historically. In such cases, it cannot be ruled out that Form *A*, representing the non-reduced form, can, apart from the primary non-reduced pronunciation, can also be pronounced the same way as Form *B*, representing the reduced form, especially in rapid, less careful speech. Examples:

(16)

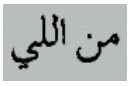


‹**ḥ**t(á)_waḥd› (*Ṣa.* 222[5])
ḥə*tta waḥəd* and *tta waḥəd*?
‘nobody’

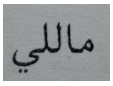


‹t(a)_waḥd› (*XM* 82[10])
tta waḥəd

(17)



‹m**n**_ally› (*ḤM* 81[8])
*mə**n**-lli* and *məlli*?
‘when’ (conjunction)



‹mally› (*ḌK* 15$_6$)
məlli

Should such a more flexible approach be adopted, *A* and *B* would be considered variational heterographs. However, in order to avoid an excessively complex description, a stricter perspective has been preferred which assumes that such graphic words are pronounceable always in accordance with their graphy.

In quite numerous pairs of graphic words, heterography is functionally ambiguous, a phenomenon responsible for the context-dependency of MA graphy. It occurs whenever at least one graphic word in a heterograph pair has ambiguous graphy. Then, determining whether these words have the same meaning (and pronunciation) is possible only on the basis of context, not the graphy itself. An example of ambiguous linear heterography is the pair composed of the univocal graphic word ما شي ‹ma_šy› *ma-ši* ‘is not’ (*Mn.* 43$_7$) and its heterograph ماشي ‹mašy›, which is an ambiguous graphy since it may have the same meaning, ‘is

not' (e.g. *Ṛḥ.* 163[11]), or a different one: 'going to' (future marker) (e.g. *MX* 21[14]) (with the pronunciations being the same). Ambiguous quantitative heterography is exemplified by the pair of the univocal graphic word كتبات ‹ktbat› *kətbat* 'she wrote' and the ambiguous graphic word كتبت ‹ktbt›, which may have the same meaning, 'she wrote', or a different one, for instance: 'I wrote' (its pronunciation being different: *ktəbt*). An example of ambiguous qualitative heterography is the pair in which گال ‹ǵal› *gal* 'he said' is a univocal graphy, while the ambiguous graphic word كال ‹kal› may have the same meaning, 'he said', or a different one, 'he ate' (with its pronunciation being different: *kal*).

Since variational heterography can be identified between words both context-dependently and context-independently (in contrast to distinctive heterography, which can only be identified irrespective of the context), some cases of ambiguous heterography are cases of variational heterography.

Two other concepts that need to be introduced at this point are expected forms and conditioned graphy. An expected form is a graphic word which has not been recorded in the corpus but its occurrence is reasonably predictable on the grounds of analogy with other forms recorded in the corpus. For instance, كارو ‹karw› for *gaṛṛu* 'cigarette' is an expected form because ك ‹k› is used to mark *g* in borrowings and, in addition, a corresponding definite noun has been found in the corpus: لكارو ‹lkarw› *l-gaṛṛu* 'cigarette^{def}' (*ḤM* 89[10]). However, positing an expected form can be risky because it may never be recorded. For instance, MA emphatic (pharyngealized) consonants which correspond to SA plain, i.e. non-emphatic (non- pharyngealized) consonants, are frequently written phonetically. Some words, however, resist this tendency, the most notable cases being the frequent words *ḍaṛ* 'house' and *ṛaṣ* 'head', which are always written دار ‹dar› and راس ‹ras›, never *ضار ‹ḍar› and *راص ‹raṣ›, respectively, in the corpus. This preference being given to donor form-oriented graphy in these particular cases can be attributed to 'the weight of the orthography of the classical language' (Aguadé 2013: 210, 214).

Conditioned graphy is a graphy which can occur only if another particular graphy is used in a given graphic word (or graphic text). Consider the following graphic word: الدواء ‹aldwao› *d-dwa* 'medicine, drug' (*Mn.* 14[11]). In order for the donor form-oriented graphy ending in ء ‹o› to be possible, the graph preceding it must be ا ‹a›, not ى ‹á›, because the latter never marks *a* word-medially (the form *الدوىء ‹aldwáo› is not expected). Conditioned graphy also affects the use of the Maghrebi traditional graphy: If a speller decides to use ف ‹f› to mark *q*, then also *f* must be marked in accordance with the Maghrebi graphy, i.e. by means of ڢ ‹f̣›, not ف ‹f›, in order to avoid ambiguity. However, using the Maghrebi ڢ ‹f̣› to mark *f* does not require marking *q* as ف ‹f›, because no ambiguity arises then (as is the case in *Ḥaš.*).

In the case of some graphic words, conditioned graphy is associated with a varying frequency of use. For example, the word *myat* 'hundred' followed by a noun denoting the counted object is most frequently written in either of the two ways: ميات ‹myat› is thoroughly self-oriented (phonetic), e.g. ميات آلف ‹myat_ãlf› *myat alaf* 'a hundred thousand' (*XF* 59[1]), while مائة ‹maỷħ› is thoroughly donor form-oriented, e.g. مائة ألف ‹maỷħ_ålf› 'idem' (*MX* 42$_{12}$). The latter form copies the SA مائة ‹maỷħ› 'hundred', in which ا ‹a› marks no sound but is a historical disambiguating graph (see point 4.a. in Sect. I.3.), ئ ‹ỷ› marks *ʔ* and ة ‹ħ› marks *t*. Forms with mixed graphy are recorded but rare: In مياة ‹myaħ›, e.g. مياة أورو ‹myaħ_åwrw› *myat oro* 'a hundred euro' (*Ḳḥ.* 126[3]), the phonetically spelt sequence ‹ya› is combined with the donor form-oriented ة ‹ħ›. In ماية ‹mayħ›, e.g. ماية ريال ‹mayħ_ryal› *myat ryal* 'a hundred rial' (*MX* 46[12]), the unpronounced SA ا ‹a› is combined with the phonetically spelt ي ‹y›. Yet other combinations, e.g. مائت ‹maỷt›, with the SA ا ‹a› and ئ ‹ỷ› combined with the self-oriented ت ‹t›, have not been recorded.

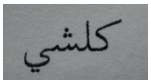
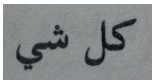
III.2.1. Variational heterography

Until the present point, examples of *minimal heterography*, i.e. one based on one distinctive unit only have been used. By contrast, *non-minimal heterography* binds heterographs which differ in more than one distinctive unit of the same kind (mixed heterography, involving distinctive units of various kinds, is dealt with in Sect. III.2.1.4.). *Maximal heterography*, in turn, means that two heterographs differ in the largest possible number of distinctive units (as evidenced by the corpus or, in a looser approach, as can be expected on the basis of analogy). Finally, there are *graphic invariants*, i.e. graphic words that have no variational heterographs at all. Identifying these would be very useful as a basis for anyone intending to propose orthographic norms for MA.

Although only qualitative heterography is dealt with in this book, some basic issues related to the two remaining types of heterography, including those introduced in the preceding paragraph, will be discussed in what follows.

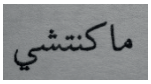
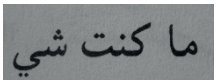
III.2.1.1. Linear variational heterography

The most frequent case of linear variational heterography is the *minimal linear variational heterography*, i.e. one binding a divisible continuous graphic word with a homosemantic and homophonous discontinuous graphic word which contains one space. In more traditional terms, this means that the same phonetic word can be written either as one word or as two words. An example, based on (7) and (10), is given in (18).

(18)  

‹klšy› (*DB* 251[1]) ‹kl_šy› (*ḤB* 35[3])
kŭll ši ‘everything’

Non-minimal variational linear heterography binds linear heterographs in which the distinctive unit consists in more than one space, as in the following example, based on (6) and (9):

(19)  

‹makntšy› (*MḤ* 45[13]) ‹ma_knt_šy› (*MX* 18_{11})
ma kŭnt-ši ‘I was not’

At the same time, (19) shows *maximal linear heterography*, because Form *A* contains the largest posssible number of spaces as distinctive units.

A category of *graphically prefixed words* needs to be distinguished in order to account for the specific behaviour of some word-medial graphies. A graphically prefixed word, or a graphic word resulting from graphic prefixation, is a linearly continuous graphic word corresponding to a phonetic word which is analyzable, whether synchronically or diachronically, into two components: the prefix (the definite article, a conjunction or a preposition) and the main component. As a result of graphic prefixation, the originally word-initial graph of the main component becomes accidentally word-medial (see Sect. I.2.). Graphs which otherwise do not occur in this position will be referred to as remnants[79]. For instance, if the definite article *l*-, written لـ ‹l› or الـ ‹al›, is prefixed to a graphic word beginning with إ ‹ą›, e.g. إبرة ‹ąbrħ› *ibra* ‘needle’, the resulting graphic word, الإبرة ‹aląbrħ› *l-ibra* ‘needle[def]’, contains word-medial إ ‹ą›, a remnant, since it does not occur word-medially in other situations. By contrast, the compositional structure of a graphically prefixed word can be obscured by writing at least one of its components in a way which is never used without graphic prefixation (such graphy may be called synthetic). For instance, if the conjunction *w*- ‘and’ is graphically prefixed to *ila* ‘if’, normally written الى ‹alá›, إلى ‹ąlá› or يلى ‹ylá›, the graphs ا ‹a›, إ ‹ą› or ي ‹y› marking the word-initial *i* may be retained:

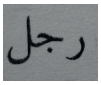
[79] This phenomenon is reminiscent of the SA spelling principle according to which graphically prefixed conjunctions, prepositions and the definite article do not influence the graphy of the initial graph of the main component.

والى ‹walá›, وإلى ‹wą́lá› or ويلى ‹wylá›, or be dropped: ولى ‹wlá› (*Bə.* 41_5). The latter graphy is an instance of synthetic graphy because *لى ‹lá› is never used for an unprefixed *ila* ‘if’.

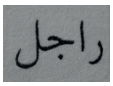
III.2.1.2. Quantitative variational heterography

Let us recall that quantitative variational heterography between two graphic words means that they are homosemantic and homophonous but one of them contains a graph or graphs absent from the other. This graph, which will be referred to as *augment*, may be added word-initially, word-finally or word-medially. In MA, it is easy to find such augments marking stable (long) vowels: For instance, in (20), the augment ا ‹a› in Form *B* is used to mark *a*.

(20)



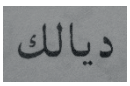
‹rğl› (*LŠ* 35_7)
ṛažəl ‘man’



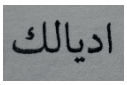
‹r**a**ğl› (*Ṣa.* 214_9)

An augment, however, may have no phonetic function, for example:

(21)



‹dyalk› (*MḤ* 69^4)
dyal-ək ‘yours’, lit. ‘of yours’

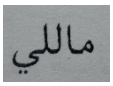


‹**a**dyalk› (*FL* 111_{10})

Here, the augment in *B* is ا ‹a› used to mark a word-initial consonant cluster.
The two above examples show pairs of quantitative heterographs differing in the absence or presence of one graph. They are thus cases of *minimal variational quantitative heterography*. Non-minimal quantitative variational heterography is easy to find. For instance, in the following pair, two augments, ا ‹a› and ل ‹l›, are added word-medially in *B*.

(22)

‹m**l**y› (*Ṣa.* 211^6)
məlli ‘when’ (conjunction)

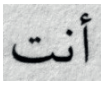


‹m**all**y› (*ḌK* 15_6)

In quite frequent cases, one heterograph has an augment (or more), while the other heterograph has another augment (or more). Put in a different way, either of the heterographs is augmented with respect to the other. This kind of quantitative variational heterography will be termed enantiodiacritic (from Greek *έναντιος* 'opposite' and *διακριτικός* 'distinguishing'). To illustrate, below are two heterographs of the pronoun *nta* 'you'. In *A*, the augment is the final ا ‹a›, while in *B*, the augment is the initial أ ‹å›.

(23)

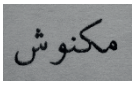
‹nt**a**› (*ḤB* 48[3])
nta 'you'



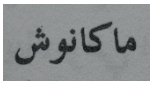
‹**å**nt› (*BB* 44[10])

In *maximal quantitative variational heterography*, one heterograph has no augments and the other has all possible augments. The pairs in (20)-(22) are instances thereof. Three others are given below.

(24)

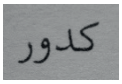


‹mknwš› (*LŠ* 35_2)
ma kanu-š 'they were not'

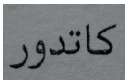


‹m**ak**anwš› (*Mu.* 5[6])

(25)

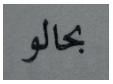


‹kdwr› (*Ṣa.* 222[2])
ka-ddur (< *ka-t-dur*) 'it[f] turns'

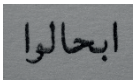


‹k**at**dwr› (ʕD 111[6])

(26)



‹bḥalw› (*ḤB* 4[9])
bḥal-u 'like him/it[m]'



‹**a**bḥalw**a**› (*XM* 57_4)

Some pairs that show minimal quantitative heterography show also maximal quantitative heterography. This is the case of (20) and (21) above. In yet another situation, some words cannot accept or lose any augments and are thus *quantita-*

tive invariants. As evidenced by the corpus, quantitative invariants include such graphic words as: باب ‹bab› *bab* ‘door’ and ديك ‹dyk› *dik* ‘cock, rooster’, i.e. those which represent phonetic words following the pattern *CVC*, with *V* being a stable (long) vowel. Further research is, however, needed in order to arrive at a complete picture of this issue.

III.2.1.3. Qualitative variational heterography

The distinctive unit in qualitative variational heterography is a graph being in graphemic variation with another graph in the same position of a graphic word. *Graphemic variation* or variation, for short, can be defined with the use of the concepts of substitutability and variational heterography. Two graphs *x* and *y* are said to be bound by variation if there is at least one graphic word in which a substitution of graph *x* for graph *y* results in its (qualitative) variational heterograph. For instance, the graph ا ‹a› is in variation with أ ‹å› because a substitution of ا ‹a› for أ ‹å› in رأس ‹rås› *ṛaṣ* ‘head’ results in its (qualitative) variational heterograph راس ‹ras›.

Depending on the relation between the sets of graphic words allowing variation, the following three types of variation are distinguished:

i. *Free variation* between *x* and *y* (symbolically: *x Fv y* or *y Fv x*) means that in every graphic word in which *x* is used it can be substituted for *y* and *vice versa*. Graphs bound by this relation are termed *free variants*.
Example: گ ‹ǵ› and ڴ ‹ĝ› in MA (in all positions[80]), because in every graphic word in which گ ‹ǵ› is used, ڴ ‹ĝ› can be substituted for it and *vice versa*.

ii. *Asymmetrical variation* between *x* and *y* (symbolically: *x Av y*) means that:
1. in every graphic word in which *y* is used, *x* can be substituted for it and
2. in at least one graphic word in which *x* is used, *y* cannot be substituted for it.
Graphs bound by this relation are termed *asymmetrical variants*. The predecessor of this relation (*x* in *x Av y*), i.e. the asymmetrical variant which can be used in a greater number of graphic words, is termed *strong variant*. Its successor (*y* in *x Av y*), i.e. the one which can be used in a lesser number of graphic words, is termed *weak variant*.

[80] For some graphs, their substitutability by other graphs depends on their position in the word. This is discussed under each pair of variants in Ch. VI.

Example: ا ‹a› *Av* أ ‹å›, word-finally, because:

1. in every graphic word in which أ ‹å› is used, ا ‹a› can be substituted for it, e.g. يبرأ ‹ybrå› and يبرا ‹ybra› *yəḅṛa* ‘he gets cured’,
2. in at least one graphic word in which ا ‹a› is used, أ ‹å› cannot be substituted for it, as illustrated by the graphic word كتبوا ‹ktbwa› *kətbu* ‘they wrote’, of which *كتبوأ ‹ktbwå› is not a variational heterograph.

Asymmetrical variation is a transitive relation. For instance, if ا ‹a› *Av* أ ‹å› word-initially and أ ‹å› *Av* آ ‹ã› word-initially, then also ا ‹a› *Av* آ ‹ã› word-initially.

iii. *Restricted variation* (symbolically: *x Rv y* or *y Rv x*) means that:

1. in at least one graphic word in which *y* is used, *x* can be substituted for it,
2. in at least one graphic word in which *y* is used, *x* cannot be substituted for it and
3. in at least one graphic word in which *x* is used, *y* cannot be substituted for it.

Graphs bound by this relation are termed *restricted variants*.

Example: ه ‹h› and و ‹w› word-finally, because:

1. in at least one graphic word in which و ‹w› is used, ه ‹h› can be substituted for it, e.g. كتبو ‹ktbw› and كتبه ‹ktbh› *kətb-u* ‘he wrote it’ are qualitative variational heterographs.
2. in at least one graphic word in which و ‹w› is used, ه ‹h› cannot be substituted for it, as illustrated in كتبو ‹ktbw› *kətbu* ‘they wrote’, of which كتبه ‹ktbh› is not a variational heterograph (it can represent, for instance, *kətb-u* ‘write it’ or ‘he wrote it’).
3. in at least one graphic word in which ه ‹h› is used, و ‹w› cannot be substituted for it, as illustrated in كلمه ‹klmh› *kəlma* ‘word’, of which كلمو ‹klmw› is not a variational heterograph (it can represent *kəllm-u* ‘he talked to him’, ‘talk to him!’ or *kəllmu* ‘they talked’).

The relations *Fv* and *Rv* are symmetrical and graphs bound by each of them can be indicated in free order, e.g. گ ‹ǵ› *Fv* ǧ ‹ĝ› is the same as ǧ ‹ĝ› *Fv* گ ‹ǵ›. By contrast, relation *Av* is not symmetrical and the graphs bound by it must be given as ordered pairs, e.g. ا ‹a› *Av* أ ‹å›, but not *أ ‹å› *Av* ا ‹a›, i.e. with the strong variant given first. Its inverse relation, symbolically *x* Av^{-1} *y*, means that *x* can be substituted for by *y* in all graphic words, but not so the opposite.

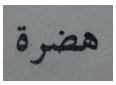
Finally, there are pairs of graphs which are never in variation, i.e. in no graphic word can graph *x* be substituted for *y* or *vice versa* without a change in meaning

and/or pronunciation. The relation binding them is *graphemic opposition*. It binds, for example, ب ‹b› and ر ‹r›, because in no graphic word can ب ‹b› be substituted for ر ‹r› or ر ‹r› can be substituted for ب ‹b› without a change in meaning and/or pronunciation. This relation is, for the time being, of no interest because it excludes variational heterography.

Below are some examples of non-minimal qualitative variational heterography, i.e. one based on the difference of more than one graph.

(27) 

‹h[a]**ḏ**ˑ**ra**› (*Ḍa.* 51[5])
həḍṛa ‘talk, conversation’

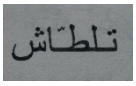


‹h**ḍ**r**ħ**› (*Ḥb* 18$_{6}$)

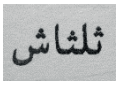
(28) 

‹a**k**t̲**r**› (*Bə.* 48[4])
ktəṛ ‘more’

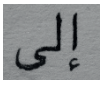
‹**å**k**t**r› (*Ṛḥ.* 163$_{1}$)

(29) 

‹**t**l̤**t̤**aš› (*Ṛḥ.* 95$_{4}$)
ṭəl̤ṭaš ‘thirteen’



‹**t̲**l**t̲**aš› (*Ḍa.* 99[13])

(30) 

‹**ą̲**l**á**› (*T̲N* 8[13])
ila ‘if’

‹**y**l**a**› (*XF* 56$_{11}$)

Maximal qualitative variational heterography means that all substitutable graphs in one graphic word are substituted for in its heterograph. Examples (27)-(30) show this relation. A special and rather exceptional case of maximal qualitative heterography is one in which all graphs in a graphic word are substitutable and are substituted for in its heterograph. The only example of this *absolute qualitative variational heterography* that I have been able to find in real texts is for SA. It concerns, rather unsurprisingly, the graphy of a short word, viz. the preposition *fī* ‘in’:

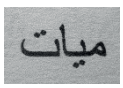
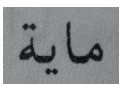
(31)  

‹**fá**› (*ʔIṯ.* 5[8]) ‹**f̣y**› (*Ḥaš.* 36)
fī ‘in’ (SA)

These two graphic words have different geographical distributions: such forms as *A* can be found, for instance, in Egyptian SA (*A* is taken from a book published in Cairo, see “ى ‹á› ~ ي ‹y›” in Sect. VI.1. on the Egyptian *yāʔ*), while *B* uses the local Maghrebi graphy and can be found in Moroccan SA. As such, they are highly unlikely to be found in a single text. In MA, the same pair of graphies can be used for the preposition *f-* ‘in’ (cognate of SA *fī*). Indeed, فى ‹fá› is used in *Ṛḥ.*, but no occurrence of في ‹f̣y› has been recorded in the corpus. Potentially, absolute qualitative variational heterographs can bind graphic words representing such phonetic words as *qfa* ‘neck’ (قفى ‹qfá› ~ قفا ‹ff̣a›), *gəffa* ‘basket’ (قفه ‹qfh› ~ قفة ‹ff̣h›), *fəqṣa* ‘sorrow, affliction’ (فقصا ‹fqṣa› ~ فقسة ‹f̣fṣa›) or *qəddəf* ‘he rowed’ (قدف ‹qdf› ~ قذف ‹fḏf̣›). The main reason why these pairs cannot been found in the corpus seems to be that the traditional Maghrebi graphy, on which they partly rely, is almost inexistent there.

By contrast, *qualitative invariants*, i.e. graphic words that admit no qualitative variation, are quite abundant, e.g. بيت ‹byt› *bit* ‘room’, كتاب ‹ktab› *ktab* ‘book’, ديال ‹dyal› *dyal* ‘of’ (possessive preposition), كنت ‹knt› *kŭnt* ‘I was’, المغرب ‹almġrb› *l-Məġṛib* ‘Morocco’. An attempt to formulate conditions for a graphic word in MA to be a qualitative invariant is undertaken in Sect. VI.2.1.

Although the greatest weight in identifying qualitative variation rests on relations between graphs rather than those between graphs and sounds, sometimes the representational aspect needs to be taken into consideration as well. This is so because the substitutability of graphs which fulfil two different functions, including marking two different sounds, should be treated in a special way. The first argument for taking the representational aspect into consideration is that without it, one type of variational heterography can be mistaken for another. Consider the second and third graphs of each graphic word in (32) (the variation of the final graphs can be disregarded here).

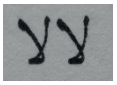
(32)  

‹m**ya**t› (*Ṛḥ.* 161_2) ‹m**ay**ħ› (*MX* 46[12])
myat ‘hundred’ (in the construct state)

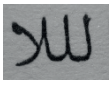
These graphs are ‹ya› and ‹ay›, respectively. From a purely formal perspective, it appears that ي ‹y› is in variation with ا ‹a›, and ا ‹a› with ي ‹y›. But when the representational functions of these graphs are considered, it becomes evident that what we are faced with here is a stable ي ‹y›, marking *y* in both heterographs, and an enantiodiacritic ا ‹a›, which in *A* marks *a*, while in *B* it marks no sound, but results from copying the SA historical disambiguating graph ا ‹a› (cf. SA مائة ‹maʾ̊ħ›). Without considering the relationships between graphs and sounds, quantitative heterography binding this pair could be mistaken for qualitative.

Another argument for taking the representational aspect into consideration is the existence of a phenomenon which will be referred to as *pseudo-variation* (symbolically: *Pv*). It occurs between two graphs which are used in two heterographic words in the same position but mark two phonetically different and unrelated phenomena. Generally speaking, this happens when in one of the heterographs, graph *x* marks sound s_1, while in the other heterograph, graph *y* marks sound s_2 (usually immediately linearly adjacent to sound s_1) or does not mark any sound at all (because it has some other function). Consider, for instance, the second graph of each graphic word in (33).

(33)

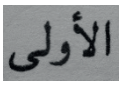
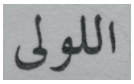


‹l**a**la› (*Fʕ* 49[8])
lalla ‘madam’



‹l**l**la› (*FL* 81[7])

These graphs, ا ‹a› and ل ‹l›, appear to be in variation. However, each of them marks a different sound: in Form *A* it is *a*, in Form *B* it is *l* (to be more precise: together with the third graph, also ل ‹l›, it marks the geminated *l*). In reality, these graphic words are bound by enantiodiacritic quantitative heterography, which is only apparently qualitative: in Form *A*, ا ‹a› is the augment, while in Form B, ل ‹l› is the augment – a situation which is not evident unless the representational criterion is resorted to. In this case of pseudo-variation, each of the two graphs, although having a different function, represents a phonetic phenomenon, viz. marks a sound. Since both of these phenomena (sounds) pertain to one variety of Arabic, viz. MA, this type will be referred to as *intra-variety pseudo-variation*. By contrast, in *inter-variety pseudo-variation*, each of the two graphs has a function belonging to the domain of a different variety. Consider the third graph of each graphic word in (34).

(34)  

‹al**å**wlá› (*Ḍa.* 66[10]) ‹al**l**wlá› (*ḤB* 12$_{7}$)
l-lŭwwla ‘first$^{f\ def}$’ (cf. SA الأولى ‹alåwlá› *ʔal-ʔūlā* ‘idem’)

These graphs, أ ‹å› in *A* and ل ‹l› in *B*, appear to be in variation. However, أ ‹å›, used in the donor-oriented graphy, reflects the sound *ʔ* pronounced in the SA cognate, while ل ‹l›, used in the self-oriented phonetic graphy, reflects the lengthened *l* in MA pronunciation. Thus, each of them carries out a different function, related to a different variety.

By contrast, graphs that have the same function but one of them is a self-oriented graphy while the other one is a donor form-oriented graphy, e.g. ت ‹t› and ث ‹ṯ›, both marking *t* in تاني ‹tany› and ثاني ‹ṯany› *tani* ‘second’, or ا ‹a› and أ ‹å›, both marking *a* in راس ‹ras› and رأس ‹rås› *ṛaṣ* ‘head’, are considered to be in variation, not pseudo-variation.

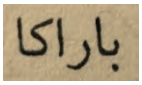
With pseudo-variation being a separate phenomenon, distinct from variation, the fact that two graphs are bound by this relation will not be considered relevant for classifying them into a single grapheme or two different graphemes (see Sect. III.2.2.2. and VI.2.3.).

Finally, it should be observed that the intensity of variation between graphs can be analyzed on two levels: abstract and actual. On the abstract level, what is of interest is the proportion of abstract graphic words in which a given type of variation occurs. On the actual level, it is the frequency of actual graphic words, i.e. graphic words recorded in the corpus, which are manifestations of a variation obtaining on the abstract level. For instance, on the abstract level, the substitution of ث ‹ṯ› for ت ‹t› by virtue of analogy (analogical graphy) can happen in *every* word containing ت ‹t›; on the level of actual words, however, it is a *marginal variation*, i.e. one represented by relatively few pairs in the corpus. Situations like this one are due to the marginal occurrence of one of the variants (e.g. ث ‹ṯ› marking *t*, ذ ‹ḏ› marking *d*, غ ‹ġ› marking *g*, ى ‹á› marking *i*) or both of them (e.g. those used for marking *ʔ* word-medially) in the corpus. By contrast, the substitutability between ه ‹h› and و ‹w› is limited on the level of abstract graphic words as it occurs only if these graphs mark the personal suffix *-u* ‘him, his; it^{m}, its^{m}’ word-finally. On the level of actual graphic words, however, it is very frequent. For a description of the MA graphy, both levels are important and this issue should be taken into consideration when proposing an inventory of MA graphemes, to which purpose, marginal variation will be considered irrelevant. Hence, any variation should also be characterized with respect to the frequency of its manifestation on the actual level.

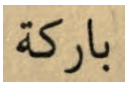
III.2.1.4. Mixed variational heterography

Variational heterography can also be based on distinctive units representing two or three different types. It is then referred to as *mixed variational heterography*. For instance, (35) shows the co-occurrence of the qualitative and quantitative type.

(35)



‹bar**a**k**a**› (*XF* 150[15])
baṛaka ‘enough’

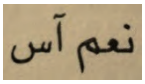


‹bark**ħ**› (*Ṣa.* 210[6])

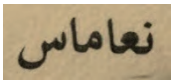
Qualitatively, the final ا ‹a› in *A* is substituted for with ة ‹ħ› in *B*. Quantitatively, *A* has the augment ا ‹a› (following ر ‹r›).

Instances of mixed variational heterography involving the qualitative, quantitative and linear type at one time are less frequent. An example thereof are the two following triply mixed heterographs representing the expression *nʕam-as* ‘sir!’ (a contraction of *nʕam, a-sid-i* ‘yes, my sir’):

(36)



‹nʕm_**ã**s› (*Bə.* 83[9])
nʕam-as ‘sir!’

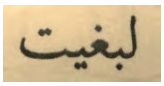


‹nʕ**a**m**a**s› (*Ḥ B* 21[4])

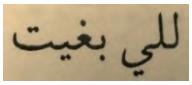
Qualitatively, آ ‹ã› in *A* is substituted for by ا ‹a› in *B*. Quantitatively, ا ‹a› is added after ع ‹ʕ› in *B*. Linearly, *A* has a space following م ‹m›, while *B* has none.

In the two above examples, the heterography of each kind was minimal. The two examples below show mixed variational heterography with the quantitative type being non-minimal.

(37)

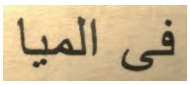
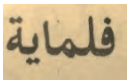


‹**l**bġyt› (*LŠ* 51_2)
lli bġit ‘what I want’



‹**lly**_bġyt› (*Ṣa.* 215_{11})

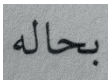
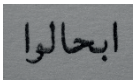
In this pair, augments ل ‹l› and ي ‹y› are added in *B* (which one of the two graphs ل ‹l› is the augment is indeterminate). The pair shows doubly mixed heterography because the graphic words also differ linearly: *B* has a space preceding ب ‹b›.

(38)  

‹f**á_a**lmy**a**› (*Ṛḥ.* 37[9]) ‹flm**ay**ħ› (*Ḥ*B 15$_{10}$)
fə-l-mya ‘per cent’

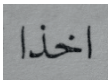
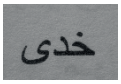
In (38), qualitatively, the final ا ‹a› in *A* is substituted for with ة ‹ħ› in *B*. Quantitatively, *A* has two augments: ى ‹á› and ا ‹a› preceding ل ‹l›, while *B* has one: ا ‹a› following م ‹m› (enantiodiacritic quantitative heterography). Linearly, *A* has a space and *B* has none. Thus, this pair shows triply mixed heterography and its quantitative heterography is non-minimal and enantiodiacritic.

Finally, *maximal mixed variational heterography* binds heterographs which are based on two or three types of variational heterography and differ in every possible distinctive unit. In other words: qualitatively, every substitutable graph is substituted for; quantitatively and linearly, every augment and space which can be added or removed is added or removed. In short: the two graphic words cannot differ more. For instance, in (39)-(41), the qualitative and quantitative variation is used up maximally (the pairs show no linear heterography since each of these graphic words is a linear invariant).

(39)  

‹bḥal**ħ**› (*BB* 107$_3$) ‹**a**bḥal**wa**› (*XM* 57$_4$)
bḥal-u ‘like him/it[m]’

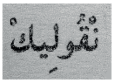
Qualitatively, ه ‹h› marking *-u* ‘him, his; it[m], its[m]’ in *A*, is substituted for by و ‹w› in *B*. Quantitatively, ا ‹a› is added word-finally (the false *ʔalif al-wiqāya*, written by virtue of analogy, see point 4.b. in Sect. I.3.) and word-initially (signalling a word-initial cluster).

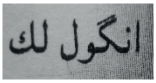
(40)  

‹a**x**ḏ**a**› (*MḤ* 66$_{11}$) ‹x**dá**› (*Ṛḥ.* 36$_{11}$)
xda ‘he took’

Qualitatively, ذ ‹ḏ› marking *d* in *A* (donor-oriented graphy) is substituted for with the phonetic د ‹d› in *B*, while the word-final ا ‹a› is substituted for by ى ‹á›. Quantitatively, ا ‹a› (signalling a word-initial cluster) is added word-initially in *A*.

Pairs showing maximal mixed variational heterography with three types of heterography are difficult to find. One is given in (41).

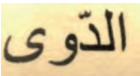
(41) 

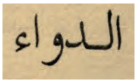


‹n˙**q**ᵘwlⁱ**y**k˙› (*Ḍ?* 13[5]) ‹**a**n**ǵ**w**l_l**k› (*XM* 119_5)
ngul-lik ‘I tell you’
(Pronunciation with *g* and *i* for *A* and *B*, respectively, is assumed)

Qualitatively, ق ‹q› in *A* is substituted for by گ ‹ǵ› in *B*. Quantitatively, *A* has the augment ي ‹y›, while *B* has a (discontinuous) gemination of ل ‹l› (which one of the two graphs ل ‹l› is the augment is indeterminate) and word-initial ا ‹a›. Linearly, *A* has no space and *B* has one.

Sometimes, the co-occurrence of quantitative and qualitative heterography involves the conditioned graphy of some graphs. Consider the following heterograph pair:

(42) 



‹ald˙w**á**› (*Ṛḥ.* 200_{10}) ‹aldw**ao**› (*Mn.* 14[11])
d-dwa ‘medicine, drug’

Quantitative variation, i.e. the addition of the augment ء ‹o› word-finally, requires a qualitative variation: ى ‹á› in *A* must be substituted for by ا ‹a› in *B*, because ى ‹á› cannot occur word-medially.

III.2.2. Grapheme

The grapheme is one of the fundamental notions in the lingustic study of writing. As the term itself suggests, its theoretical status within graphemics is assumed to be analogous to that of phoneme in phonology. This analogy is true at least insofar as the grapheme, like the phoneme, has been conceptualized and defined in many various ways since this term was used for the first time by Baudouin de Courtenay (1901: 116, cf. Ruszkiewicz (1981b [1978]: 21, 27). For the needs of the present work, instead of recapitulating its history (for an overview see Kohrt 1985), it seems preferrable to present the two principal conceptions that have evolved around it and which will be referred to, following Günther (1988: 72), as

the *representational conception* ('Repräsentanzkonzeption') and the *distinctivity conception* ('Distinktivitätskonzeption'). In the following paragraphs, they will be presented and discussed in the context of MA graphy. Further on, problems related to describing MA graphy with the use of the concept of grapheme within the distinctivity conception will be highlighted and a modified approach proposed.

III.2.2.1. Two main approaches to the concept of grapheme

According to the representational conception, a grapheme is understood as a graphic representation of a sound or a phoneme. Graphemicists who adhere to this view ask questions about how particular sounds or phonemes can be represented in writing by means of graphemes. The result of their analyses are lists of correspondences between sounds, or phonemes, and graphemes. As far as linguists are concerned, this view is held, for instance, by Hammarström, who defines the grapheme as "the class of graphs which denote the same phoneme" (1981 [1964]: 97). It can also be encountered in publications intended for a wider audience. For example, the *Oxford English Dictionary* defines a grapheme as:

> The class of letters and other visual symbols that represent a phoneme or cluster of phonemes, as e.g. the grapheme ‹f› consists of the *allographs* (…) *f*, *ff*, *F*, *Ff*, *gh*, *ph*, and *Ph* which represent the phoneme /f/ in *fun*, *huffy*, *Fingal*, *Ffoulkes*, *cough*, *graph*, and *Philip* respectively; (…) ("Grapheme, *n.*", 2018).

In contrast, the distinctivity conception is based on an assumption that writing is a self-contained system and, consequently, it regards the grapheme as a language unit independent from whatever sound(s) or phoneme(s) it can represent. Within this approach, a grapheme is conceived of as a unit that should be analyzed with respect to other graphemes in terms of its distinctive functions, without asking questions about its representational function, which matters on a different level, the phonetic one. In this manner, the grapheme is defined by Günther (1988: 72) as the least meaning-differentiating unit of written language. This is where its resemblance to the phoneme is believed to reside[81].

The feasibility of establishing graphemes in a way analogous to phonemes has been challenged by some linguists. For instance, Kohrt argues that the grapheme and the phoneme "can never be really on a par" (1986: 91), the main reason for this being that they have different domains: The written units (graphs, letters,

[81] Because of this postulated analogy between the grapheme and the phoneme, Kohrt (1986: 84) terms this conception *analogical conception* (or *conception of methodological analogy*). *Reference conception* is his term for Günther's representational conception.

There are also other uses of the term 'grapheme', one of them being a synonym for a 'letter' or other unit of a script, "with no theoretical content at all" (Daniels 2013: 415, footnote 9).

graphemes) are pre-established for a specific purpose, or "prestabilized from the outset". Only at a later stage are they 'blended', resulting in a continuum, which is thus secondary. In spoken language, by contrast, the continuum is primary, and its units must be detected and individuated (p. 90). Besides, as Kohrt observes, "[w]riting always presumes some kind of consciousness (which is not necessary in the sphere of spoken language)" (p. 93), with which Daniels (2013: 415) agrees, speaking of writing as "the consciously devised phenomenon". Another reason why the number of analogies between the units and relations of written language and those of spoken language is limited is because the influence and impact of an individual user on the totality of written texts (relatively smaller) are much stronger than those on the totality of spoken utterances (relatively larger). This is one of the reasons why, as Günther (1988: 79) points out, the relationships between units of written language are much more unsystematic than those between units of spoken language (cf. also Rogers 2005: 11, who remarks that "writing is much more varied in its structure than speech"). For a discussion of the controversy on the possibility of a structural graphemics, see Daniels (1992, 1995) and Herrick (1995a, 1995b). In the following section, the problem of the grapheme will be discussed in the context of MA against the background of its phonological relations.

III.2.2.2. The grapheme in the context of Arabic script

Previous research on the Arabic script used for MA, and SA as well, which was presented in Sect. II.1., can be characterized as following the representational approach: It concentrates on describing which letters or signs (sometimes referred to as 'graphemes', e.g. Aguadé 2006: 256, footnote 18, 259) are used to mark particular sounds. However, should the representational conception of the grapheme be consistently applied to MA, it would have to be specially modified in view of some peculiarities of its graphy. One of them is that two classes of graphemes would have to be distinguished: obligatory graphemes and facultative graphemes, corresponding to the primary and secondary graphic system, respectively, discussed in Sect. I.2. Secondly, and more importantly, following the model of SA, the Arabic script used for MA uses graphs which do not represent any sounds, but have other functions: For instance, the graph ا ‹a› marks some verbal categories (the *ʔalif al-wiqāya*) or signals a word-initial consonant cluster (pseudoprothesis). On the other hand, some MA sounds, in particular vowels, may be not represented in script. Thus, when drawing up an inventory of MA graphemes understood representationally, one would have to depart from two different points: from graphemes, to which sounds are assigned, and from sounds, to which graphemes are assigned. Another difficulty in this approach is that graphic words are frequently phonetically ambiguous (see Sect. III.1.), which sometimes makes as-

sociating graphemes with sounds (or phonemes) very unsystematic. This applies to both MA and SA. For instance, in SA, ة ‹ħ› marks *t* in graphic words used in the construct state, but in pausal positions, it is interpretable as marking either *a* or no sound at all (if the task of marking *a* is assumed to be carried out by the *fatḥa*, a vocalization mark hardly ever written in this position). In the free state, this graph marks either *a* or *t*, often depending on the speaker's free choice. The representational function of ة ‹ħ› also depends on semantic factors: For instance, it is practically never pronounced when used in graphic words representing personal names. Within the representational approach, special descriptive treatment would also be required for MA words following a donor-oriented graphy which is at odds with MA pronunciation: This approach would be at a loss as to how to reflect the fact that in the case of the donor-oriented graphy the relationships are not 'MA meaning > MA phonetic form > MA graphic form', but rather 'MA meaning > SA phonetic form > MA graphic form'.

The distinctivity approach, by contrast, does more justice to the self-containedness of the writing system. This relatively strong independence of written language from spoken language is especially visible in the case of Arabic script. While reading texts written in this writing system, it is usually not necessary to know the pronunciation of particular words, especially as far as vowels are concerned (let alone to read these words aloud) in order to understand what they mean[82]. Thus, in reading a MA text written in Arabic script, knowing the meaning is, in a way, cognitively prior to knowing the pronunciation.

However, while this study gives precedence to the distinctivity approach in order to focus on variational heterography with the aim to establish graphemic relations between graphs and thus arrive at graphemes, the representational conception is sometimes resorted to as well. Representational relationships between graphs and sounds cannot be disregarded completely, unless one is willing to negate the fact that writing has been developed as a means of reflecting spoken language. In numerous cases, a close look at representational relationships helps to identify spelling principles underlying particular graphies, especially to distinguish variation (different spellings for the same phonetic phenomenon) from pseudo-variation (different spellings signalling different phonetic phenomena).

The distinctivity approach is not without its problems, not only as far as the MA graphy is concerned. As observed above, relationships between the units of written language have been characterized as less regular in comparison with those binding the units of spoken language. However, even the latter are not always

[82] Cf. Abu-Rabia & Taha (2016: 325-326), who observe: "The reading-aloud tasks require phonology and grapheme-phoneme conversion. Silent-reading comprehension, however, may require visual-orthographic identification of roots of words for initial lexical access more than phonological representation".

fully systematic: There are cases in which a pair of sounds is capable of differentiating meaning context-independently in some contexts, while in others, they are context-independently mutually interchangeable. In yet other word pairs, their relation is context-dependent, i.e. context is needed to determine whether two words differentiated only by the opposition of these two sounds have the same meaning or not. In other words, their phonetic opposition is functionally deficient. Two pairs of MA sounds, *s* ~ *š* and *q* ~ *g*, will be used to illustrate this. The pair *s* ~ *š* is a clear context-free case of phonological opposition if, for instance, the pair *sqa* 'he gave (someone) to drink; he fetched (water)' vs. *šqa* 'he got tired' is considered. But in some other pairs, these sounds are in free variation: for example, 'sun' in MA is both *šəms* and *šəmš*, and 'window' is both *səṛžəm* and *šəṛžəm*. This pair of sounds can also be in a neutralized relation, i.e. it is only from context that one can determine the semantic relationship between *šədd* and *sədd*: whether they mean the same, i.e. 'he closed', or two different things, for instance 'he closed' vs. 'a dam'. The pair *q* ~ *g* is similar in this respect. Used in some words, these sounds are capable of differentiating meaning irrespective of context, e.g. *gana* 'good mood' vs. *qana* 'ditch', and thus should be considered to be bound by phonological opposition. But in other words, they are in variation: 'he said' is both *gal* and *qal*, and 'in front of' is both *gŭddam* and *qŭddam*. There are also word pairs the semantic relationship between which can be determined only from the context, for instance, *gaʕ* and *qaʕ* can mean the same, 'bottom, pedestal' or two different things: 'all' and 'bottom, pedestal', respectively.

Perhaps the relative numbers of word pairs representing particular relations could be resorted to in order to determine the functional status of such pairs of sounds. In MA, some pairs of sounds predominantly differentiate meaning context-independently, while words in which they are in variation, or in a neutralized relation, are marginal. This is the case of *s* and *š*. Their predominant meaning-differentiating capability is an argument for classifying them into two phonemes. Some other sounds, by contrast, are predominantly in variation, while pairs in which they differentiate meanings context-independently are rare. The pair *g* and *q* is an example of this. Their marginal meaning-differentiating capability might be an argument for placing these two sounds in one phoneme, as variants[83]. However, deciding whether the meaning-differentiating capability is strong or marginal can be difficult for some pairs, e.g. *s* and *ṣ*, without a statistical count.

[83] Relationships such as the one between *q* and *g* have turned out problematic for scholars. For instance, Sánchez (2014: 93) says that (in the dialect of Marrakesh) /g/ is considered an allophone of /q/ 'in some words in which it is a free variant', but in the next sentence he adduces pairs which 'fully justify the phonemic status of both realizations'. This can only mean that *g* and *q* are both in one phoneme and in two different phonemes at the same time. Laconically, Danecki writes of 'an imbalance between the phonemes *q* and *g*' (1989: 141).

Instances of the unsystematicity of functional relationships between units can be found even more easily in written language. In the MA graphy, it frequently happens that a pair of graphs is capable of differentiating meanings context-independently in some pairs of graphic words, while in others, they are freely interchangeable irrespective of the context (graphemic variation), and in yet others their graphemic relation is neutralized, i.e. it is context that determines the semantic relationship between the graphic words (the latter situation occurs if at least one of the heterographs is an ambiguous graphy). Consider the following examples of pairs of graphs and functional relations which can bind them (phonetic relationships do not concern us at this point).

Example 1: ق ‹q› and ك ‹k›, in all positions

a. Meaning-differentiating function, e.g. in the pair: قرا ‹qra› *qṛa* 'he read, he studied' vs. كرا ‹kra› *kra* 'he rented';

b. Graphemic variation, e.g. in the pair: قدام ‹qdam› vs. كدام ‹kdam› *gŭddam* 'in front of';

c. Neutralized relation, e.g. in the pair قال ‹qal› vs. كال ‹kal›, which depending on the context may have the same meaning and pronunciation: *gal* 'he said', or two different ones: *gal* 'he said' vs. *kal* 'he ate'.

Example 2: ه ‹h› and و ‹w›, word-finally

a. Meaning-differentiating function, e.g. in the pair: كراه ‹krah› *kra-h* 'he rented it' vs. كراو ‹kraw› *kraw* 'they rented';

b. Graphemic variation, e.g. in the pair: مدرسته ‹mdrsth› vs. مدرستو ‹mdrstw› *məḍṛaṣṭ-u* 'his school';

c. Neutralized relation, e.g. in the in the pair: كتبه ‹ktbh› and كتبو ‹ktbw›, which depending on the context may have the same meaning and pronunciation: *kətb-u* 'he wrote it', or two different meanings (and pronunciations): كتبه *kətb-u* 'he wrote it' or *kətba* 'a writing' vs. كتبو *kətbu* 'they wrote', 'write!^{pl}' or *kətb-u* 'write it!'.

Relatively numerous pairs of graphs are bound in MA graphy by what is referred to here as the neutralized relation. In many cases, for example that of ق ‹q› and ك ‹k›, it is difficult to tell if they are used more in the meaning-differentiating function or in graphemic variation, in contrast to pairs for which one relationship is dominant. Therefore, for the purpose of proposing an inventory of MA graphemes, the different types of qualitative variation will be considered and used to negotiate descriptively the problem of these functional multivalencies. The priority in this approach to classifying graphs into graphemes is given to variation rather than opposition. Thus, the criterion proposed provides: *Two graphs are classified*

into one grapheme if and only if they are bound by free variation or asymmetrical variation, in all possible positions, and the variation is not marginal.

This criterion has the following consequences: Two graphs that are never in variation with each other are classified into two different graphemes. A graph that has no variants at all is classified into a separate grapheme. Two graphs that are in restricted variation or pseudo-variation can be in one grapheme but neither of these relations entails classifying the pair into one grapheme.

Variants belonging to one grapheme are called *allographs*. An allograph which occurs in restricted environments, i.e. depends on some additional conditions, is a *non-basic allograph*. The allograph which is independent of such additional conditions and can be used in a greater number of environments is the *basic allograph*.

Before the issue of graphemes in MA is taken up again in Sect. VI.2.3., particular cases of variation between graphs will be illustrated and analyzed in Sect. VI.1. This will be preceded by a discussion of some other specific problems related to the description of MA graphy and of the spelling principles operating in it.

IV. Moroccan Arabic graphy. Descriptive problems

In the following paragraphs, some methodological problems related to the description of MA spelling are discussed (some of them are indicated in Michalski 2017). First, this concerns various types of indeterminacy of graphs and graphic words, i.e. the impossibility to identify in a non-arbitrary way a unit under description or relations that bind it with other units. Second, the issue of idiosyncratic and rare graphies and their descriptive treatment will be discussed.

IV.1. Graphic indeterminacy

Graphic indeterminacy can be explained with the use of the concepts of actual graphs and abstract graphs (see Sect. III.1.). It consists in the impossibility of associating an actual graph or sign used in a given text with one of the abstract graphs existing in the language awareness of the language users. In plain words, this means that what is written, printed or displayed in a particular place of a particular text cannot be deciphered. Consider, for instance, (1) and (2).

(1) 

‹**å**$^{\mathbf{u}}$hw› or ‹**å**hw›? (*Ḍa.* 65$_{10}$)
u-hŭwwa ‘and he’

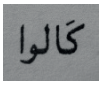
(2) 

‹**å**$^{\mathbf{a}}$d$^{\cdot a}$ah› or ‹**å**d$^{\cdot a}$ah›? (*MX* 33$_{2}$)
dda-h ‘[it] took him’

The first graph in each of these two graphic words is clearly أ ‹å›. However, the *hamza* sign overlaps in each case with an indistinct sign, probably a *ḍamma*, ـُ ‹^{u}›, in (1) and a *fatḥa*, ـَ ‹^{a}›, in (2) (note that the latter is normally written above, not below, the *hamza*). Such situations are often in both sources. (3) and (4) are other examples of graphic indeterminacy.

(3) 

‹**m**$^{\cdot}$h^{a}a› or ‹**m**$^{\bullet}$h^{a}a›? (*Mm.* 112^{2})
ṃṃa-ha ‘her mother’

(4) 

‹**ǵ**alwa› or ‹**k**$^{\mathbf{a}}$alwa›? (*XM* 70$_{7}$)
galu ‘they said’:

In (3), a partly blackened circle appears as a secondary graph (a situation quite often in *Mm.*). It can only be guessed that this represents a *šadda*, ـّ ‹ˑ›, superposed on a *sukūn*, ـْ ‹ˑ›. In (4), the first letter has a short stroke above which is printed in a way that makes it look like a *fatḥa*, ـَ ‹ᵃ›, a situation quite often in *XM*. However, after comparing a number of words containing this graph in this source one can be sure that the stroke is not a *fatḥa* but a constituting part of the graph گ ‹ġ›.

These and similar problems, resulting from the imperfections of printing technology, are rather rare and can be neglected. Examples affected with this kind of indeterminacy are not used in the work.

IV.2. Functional indeterminacy

Functional indeterminacy means the impossibility of determining the function of a given graph. A special case thereof is phonetic indeterminacy, which means that it is known that the function is marking a sound but it cannot be determined which one (see Sect. IV.2.1.). First, however, two other cases of functional indeterminacy will be discussed: one concerning some graphs used word-initially and another one concerning some graphs used word-medially at the boundary of two morphemes.

The first example concerns the function of أ ‹å› used word-initially, e.g. in أحكى ‹åḫká› (*Ḍa.* 75[6]). This graph can be interpreted either as a pseudo-prothesis indicating a word-initial consonant cluster; the pronunciation is then *ḥka* and the meaning ‘he told’. Another interpretation is that it marks the conjunction *u-* ‘and’, with the pronunciation being *u-ḥka* and the meaning ‘and he told’ (on this practice, see Aguadé 2006: 267-268). The latter interpretation is suggested by context, by other graphic words which more clearly show that this conjunction is written in this way and by the fact that *ḥka* is written with no pseudo-prothesis in this source, e.g. الصغيرحكى ‹alṣġyrḥká› *ṣ-ṣġiṛ ḥka* ‘the little one told’ (*Ḍa.* 75[2]) (the absence of space is irrelevant here). What goes against this interpretation is that other graphic words can be found in *Ḍa.* in which أ ‹å› is clearly used as a pseudo-prothesis, e.g. in أبدت ‹åbdt› in the phrase: أمنين أبدت ‹åmnyn_**å**bdt› *u-mnayn bdat* (or *mnayn bdat*) ‘and when she began’ (or ‘when she began’) (*Ḍa.* 48[7]). Another example is that of the graphic word اشنو ‹ašnw› ‘what’ (*ṮN* 46[11]), which has two possible pronunciations: *šnu* and *ašnu*. If the former pronunciation is assumed, the function of ا ‹a› is to signal a pseudo-prothesis. If the pronunciation is taken to be *ašnu*, the graph marks the vowel *a*.

The second case of functional indeterminacy concerns some graphs used word-medially. For instance, it cannot be determined whether ا ‹a› in كادّوب ‹kadˑwb› *ka-ddub* ‘you melt’ (*Ṛḥ.* 13[5]) is used to mark the vowel *a* in the preverb

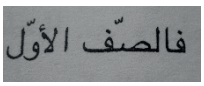
ka- or is a remnant of the word-initial pseudo-prothesis signalling the gemination *dd*. Either interpretation can be supported by the speller's decisions found elsewhere in his text: On the one hand, *ka-* is normally written كا ‹ka› there. On the other hand, a pseudo-prothetic ا ‹a› is used in ادّوب ‹ad·wb› *ddub* 'she/it^{f} melts' (*Ṛḥ.* 191[3]).

The use of the above or similar cases will be avoided in the subsequent analysis and clearer examples will be adduced instead.

IV.2.1 Phonetic indeterminacy

Phonetic indeterminacy is a phenomenon situated at the intersection of functional indeterminacy and phonetic ambiguity (see Sect. III.1.). It occurs when a graphic word or graph has more than one pronunciation, i.e. it is phonetically ambiguous, and the pronunciation intended by the speller cannot be determined even when the meaning is known and a context given. The difference between phonetic ambiguity and phonetic indeterminacy consists in that the former merely denotes the fact of having more than one pronunciation, irrespective of whether the pronunciation intended by the speller can or cannot be identified by the reader, e.g. with the help of the context, while the latter concerns cases where the identification of the intended pronunciation is impossible.

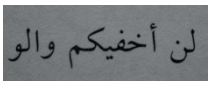
The basic and most common case of phonetic indeterminacy concerns two (or more) possible pronunciations which can be associated with a graphic word. This can happen between a MA graphic word and its SA cognate or between two MA words. The former situation means that it is impossible to determine whether a graphic word represents a MA word or its SA cognate. It occurs if the MA word is phonetically similar to its SA cognate or is written in a donor form-oriented way. An example of this is the graphic word الأوّل ‹alåw·l› 'first^{def}' in (5).

(5) 
فالصّف الأوّل

‹falṣ·f_alåw·l› (*ṬN* 44[4])
'in the first row'
fə ṣ-ṣəff l-lŭwwəl or *fə ṣ-ṣaff^{f} l-ʔawwal* ?
(cf. SA في الصّفّ الأوّل ‹fy_alṣ·f·_alåw·l› *fī ṣ-ṣaff^{f} l-ʔawwal^{i}* 'idem')

Although its meaning is known, its pronunciation is indeterminate: it can be either MA *l-lŭwwəl* (cf. the univocal graphy اللوّل ‹allw·l›, used in the same phrase in *Ṛḥ.* 30[2]) or SA *l-ʔawwal*. Since the context in which the entire phrase is used concerns religious matters, SA pronunciation may be intended.

This type of indeterminacy would not occur if one could be sure that a text under analysis is consistently written in one particular variety of Arabic: MA or SA. However, linguistic interference, or code-mixing, between MA and SA is encountered in both spoken (cf. Durand 2004: 40-42) and written contexts[84], to which the following example bears witness:

(6) 

‹ln_åxfykm_walw› (*BB* 160_2)
lan ʔuxfiya-kum walu
'I will not keep anything back from you^{pl}'

In this phrase, the negated verb with the suffixed pronoun, meaning 'I will not keep back from you^{pl}' has the SA future negation particle *lan* and the SA verbal affix *ʔu-* (however, the rection of the verb is not correct SA: It should be *ʔuxfiya ʕalay-kum*, i.e. with the preposition *ʕalā*). In contrast, the word *walu* 'nothing' is clearly MA. Such cases of code-mixing suggest that authors can be expected to use SA words in phrases which appear to be written entirely in MA. For the needs of the present work, it is assumed that if a graphic word used in a MA context has a SA form but is pronounceable in the MA manner, it represents a MA word spelt in the donor form-oriented graphy.

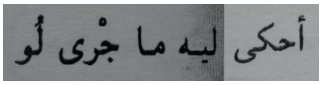
As for phonetic indeterminacy between two MA words, these two words may belong to two different varieties of MA or to a single one. Sometimes, it is difficult to determine the variety of MA a text is written in or the pronunciation intended by the speller (cf. Aguadé 2005: 247, 2.2.1.). A good illustration of this problem concerns the pronunciation of the expression resulting from attaching some personal suffixes to the preposition *l-* 'for, to'. The resulting form may contain *-i-*, e.g. *li-k* 'to you' and *li-h* 'to him', or not, *lə-k* 'to you' and *l-u* 'to him'. Aguadé characterizes the pronunciation of this expression as "[d]epending on the dialect" (for instance, "[p]rehilalian dialects have always forms without it" [i.e. without *-i*], Aguadé 2006: 267). Some graphies are unambiguous in this respect: The graphy with ي ‹i› always stands for forms with *i*: ليه ‹lyh› clearly represents *li-h*, while that

[84] Durand (2004: 40) actually equals what results from code-mixing between vernacular MA and SA to the variety which is intermediate between them, termed 'middle Arabic', *al-ʕarabiyya al-wusṭā* (or 'modern MA' in Youssi 1992).

For a discussion of linguistic interference illustrated with EA and Syro-Lebanese Arabic material, see Diem (2006: 25-52, esp. 36-52 for the morphological, syntactic and lexical levels). Cf. also Mejdell (2014), dealing with strategic bivalency, i.e. elements belonging to the two different varieties of Arabic deliberately used and manipulated by language users, in EA.

with و ‹u› always stands for forms with *u*: لو ‹lw› clearly represents *l-u*. However, the graphy with no ي ‹i› or و ‹u›, له ‹lh›, is ambiguous: it can be read both ways: *li-h* or *l-u*. The form and meaning of the expression resulting from attaching suffixal pronouns to this preposition have been presented in various ways in descriptions of MA[85], which makes the description of the graphic words representing them even more problematic.

This problem also affects texts written in a single variety and even those representing the speech of a single character of a narrative. For instance, on one page in *BB*, when one of the characters of the novel says 'tell them!', her words are represented by means of the ambiguous graphy گُلْ لهم ‹ǵulˑ_lhm› (*BB* 213[1]) in one line, while in another the univocal one گُلْ ليهمْ ‹ǵulˑ_lyhmˑ› (*BB* 213[4]) is used. Hence, it might seem that these two graphies both represent the same pronunciation and that the ambiguous graphy should be read *gul li-hŭm* rather than *gul l-hŭm*. However, it can also happen that two different univocal graphies indicate the use of the two paradigms in a text or a line representing the speech of a single character. For instance, forms belonging to the two different paradigms, *li-h* and *l-u*, are used in the narrator's words in *Ḍa.* in one line:

(7) 

‹åḥká_**lyh**_ma_ǧˑrá_**lᵘw** › (*Ḍa.* 75[6-7])
u-ḥka li-h ma žra l-u
'and he told him what happened to him'

This and similar cases (which can also be found, for instance, in *ḤM*) suggest that the use of these two paradigms by individuals may be inconsistent.

Another instance of phonetic indeterminacy within a single dialect, one of the quantitative type, too, concerns the initial ا ‹a› in some graphic words. For instance, ا ‹a› in اتكتبوا ‹atktbwa› *tkətbu* 'you[pl] write', could be best interpreted as

[85] Some authors give only one paradigm: Thus, only forms without *-i-* are given in Harrell's grammar (1962: 138, 143) and used throughout his dictionary (Harrell 1966), both works being based on the speech of the educated urban speakers from Fez, Rabat and Casablanca. By contrast, forms with *-i-* are given in Aguadé & Benyahia's dictionary (2005, in particular p. 87, s.v. *l-*), based on the Casablanca variety, and Sánchez (2014: 218-219, 234), a description of the dialect of Marrakesh. According to Caubet (1993: 206), there are two prepositions that are different semantically: *l-* (the short form) "marque un déplacement vers quelque chose" and "généralement la construction indirecte" (e.g. in *šnu gal lə-k?* 'What did he tell you?'), while *li-* (the long form) "marque l'attribution" (e.g. *ma-ši li-ya* 'is not mine') or "le destinataire" (e.g. *šrit-u li-k* 'I bought it for you'). Durand (2004: 90) observes that the paradigm of the preposition *l-* with suffixed pronouns (which he calls *serie indiretta*) is split ("sdoppiata"), for which, however, he gives no further explanation.

signalling an initial consonant cluster. However, as Durand points out, the pronunciation *tkətbu* (*tkətbu* in his notation), although prevalent, is not the only one possible: One can also hear: *təkətbu*, and *təktəbu* (Durand 2004: 128), i.e. forms with no initial cluster, which would make the graphy with the pseudo-prothetic ا ‹a› unjustified.

The use of the graph ق ‹q› will be discussed now to illustrate phonetic indeterminacy affecting qualitative features. In MA graphy, this graph is phonetically ambiguous as it is used to mark two sounds: *q* or *g*. In some cases, the graphic word itself or the context suffice to determine which one of them is to be pronounced. For instance, it can be only *q* in قليل ‹qlyl› *qlil* 'little, few' and only *g* in فازق ‹fazq› *fazəg* 'wet'. In some contexts, however, it is phonetically indeterminate: For instance قال ‹qal› 'he said' can be pronounced *qal* or *gal*, with the pronunciation depending on the variety but also the speaker[86]. However, like in the case of the preposition *l-* with suffixed pronouns, different graphies can be used for representing the speech of a single person. For instance, when the narrator and protagonist in *ḤB* says 'he said' in MA, it is written, on a single page, by means of the ambiguous graphy قال ‹qal›, representing either *qal* or *gal* (*ḤB* 4[4]), and the univocal graphy گال ‹ĝal›, representing *gal* (*ḤB* 4[6]). How should such situations be interpreted? Arguably, even if it cannot be excluded that one person shifts from one pronunciation to another, it is rather unlikely for an author to reflect this deliberately in writing. Hence, for the purpose of the present description it is assumed that in cases like this the pronunciation is the same, viz. that قال ‹qal› and گال ‹ĝal› form a variational heterograph pair. However, the possibility that an author deliberately writes ق ‹q› to mark *q* should always be taken into consideration.

Finally, an issue to be discussed here is the parallelism of double pronunciation and double graphy. By this term, phonetic variation accompanied by a parallel heterography is meant. For instance, MA has two phonetic words for 'he managed': *dəbbəṛ* and *ḍəḍḍəṛ*. Of interest to us is the variation between the sounds *d* and *ḍ* (the remaining varying sounds are irrelevant). Parallelly to this double pronunciation, two graphies are in use: دبر ‹dbr› and ضبر ‹ḍbr›. This situation can be interpreted in two ways. One is to assume that there is strict correspondence between graphy and pronunciation, viz. that ض ‹ḍ› marks *ḍ* and د ‹d› marks *d*. The other possiblity is to assume that in view of the fluctuation in pronunciation and spelling, such a correspondence does not exist. Accordingly, the graphic word دبر ‹dbr›, apart from its more literal pronunciation *dəbbər*, could also be pronounced *ḍəḍḍəṛ*, while the graphic word ضبر ‹ḍbr›, apart from its more literal pronuncia-

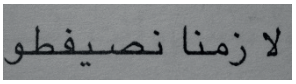
[86] Knowing the origin of the author of a given text is helpful only to a limited extent. For instance, even if it is known that the author uses the *koine* based on the Casablanca variety, still both *q* and *g* are possible (cf. Aguadé & Benyahia 2005: 9, e; Ennaji 2013: 238. For a discussion of the variation between *q* and *g* in MA in general, see Heath 2002: 141-147).

tion *ḍəḅḅəṛ*, could also be pronounced *dəbbər*[87]. The latter interpretation, although appearing to adequately reflect the real usage, would require a much more complex and speculative description, with too much space being left for uncertainty. Therefore, the first, simpler option has been chosen in this work.

IV.3. Idiosyncratic graphies, misprints and rare graphies

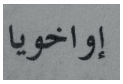
Some graphies belong to the idiosyncratic usage of a single author only. This is, for instance, the case of the use of *matres lectionis* for marking reduced vowels, e.g. in كونت ‹kwnt› *kŭnt* ‘I was’ (*Ṛḥ.*, passim), which has been found only in *Ṛḥ.* Despite their idiosyncratic use, they are included into the macrosystemic description as long as such a use can be considered a tendency in a given source. Such graphies are treated descriptively as equally valid to more frequent ones, but their individual character is indicated.

Graphies that seem to be evident misprints are not taken into consideration. This decision might seem problematic considering the fact that with no orthographic norms, one cannot speak of errors, i.e. graphies which break such norms. A misprint, however, can be identified as something different from an error. It is understood as a graphy used only once for which no other explanation can be found than a technical fault or misunderstanding of the text on the part of the ultimate creator of a given graphy, i.e. the typesetter. The example in (8) concerns linear features and shows a misprint consisting in placing an unnecessary space after لا ‹la›, perhaps because it was mistaken for *la* ‘no, not’.

(8) 

‹la_zmna_nṣyfṭw› (*ṮN* 52_{12}) (instead of لازمنا نصيفطو ‹lazmna_nṣyfṭw›)
lazəm-na nṣifṭu ‘we have to send’

The following example, by contrast, shows a missing space:

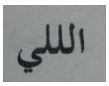
(9) 

‹ąwaxwya› (*ḤB* 14_{12}) (instead of إوا خويا ‹ąwa_xwya›)
ʔiwa xu-ya ‘well, brother’

---

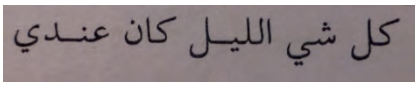
[87] Other examples of this parallelism are: شمس ‹šms› vs. شمش ‹šmš› ~ *šəms* vs. *šəmš* ‘sun’; جوج ‹žwž› vs. زوج ‹zwǧ› ~ *žuž* vs. *zuž* ‘two’; شرجم ‹šrǧm› vs. سرجم ‹srǧm› ~ *šəṛžəm* vs. *səṛžəm* ‘window’.

A quantitative misprint is shown in the graphic word in (10), which contains at least one ل ‹l› too much.

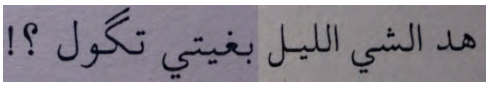
(10) 

‹allly› (*Mu.* 44[9]) (instead of اللي ‹ally›, للي ‹lly›, لي ‹ly› and other forms)
lli ‘which’

Even though repeated, some forms should be considered misprints if only a repeated typographical error can be their justification. In the following example of a repeated quantitative misprint, *lli* ‘which’ is misprinted as الليل ‹allyl› *l-lil* ‘night’, a word which makes no sense here:

(11) 

‹kl_šy_allyl_kan_ʕndy› (*TM* 67$_{10}$)
kŭll ši lli kan ʕənd-i
‘all that I had’

(12) 

‹hd_alšy_allyl bġyty tġwl› (*TM* 68[7-8])
had š-ši lli bġiti dgul
‘the thing you wanted to say’

Some graphies, however, are dubious in this respect, i.e. it is unclear whether they are misprints or a tendency because they appear repeatedly and may have some kind of justificaiton. A conspicuous example concerns a qualitative feature, namely the use of ي ‹y› word-finally to mark *a*, while one could think that ى ‹á› is intended. For instance, the word *mša* ‘he/it^{m} went’ is sometimes written مشي ‹mšy› (*Ṛḥ.* 168$_{9}$) instead of the current graphy مشى ‹mšá› and the word *ila* ‘if’ is written يلي ‹yly› (*TN* 53[3]) instead of the more current graphy يلى ‹ylá›. This phenomenon is repeated (although in our corpus it occurs mostly in *Ṛḥ.*) and, in addition, can be justified in some way: First, since it also appears in SA, it may be interpreted as being modelled on it. Second, it can be interpreted as an instance of analogical graphy (see the next chapter). For these reasons, its instances are not rejected as misprints but included in the description.

V. Spelling principles in Moroccan Arabic

In this chapter, the issue of spelling principles presented introductorily in Sect. II.1. is taken up again. Here, they will be supplemented, systematized and refined. The task of identifying a spelling principle underlying a particular graphy may, however, be not easy since the spellers of MA practically never indicate reasons for the way they write.

Due to their non-normative character, the principles are referred to as spelling principles, not orthographic ones. Also, since it is possible that such principles can also be identified in other alphabetical or semi-alphabetical writing systems borrowed from (genetically related) donor languages, their names are kept as little language-specific as possible.

Based on the fundamental decision a speller writing in MA has to make, i.e. one concerning orientation, i.e. what is used as the point of reference of the MA graphy, two major principles were distinguished: the principle of self-orientation (the point of reference being MA) and the principle of donor-orientation (the point of reference being SA, the graphy of which is borrowed or imitated). The point of reference of a third principle, that *of donor-defying*, is SA, as well, but treated negatively: It means that although the SA graphy would well reflect MA pronunciation (or morphology), the speller rejects it in favour of some other (phonetic or morphological) graphy. For this reason, it is treated here together with the self-orientation principle.

Each of these two major spelling principles divides into two further principles.

The principle of donor-orientation divides into the principle of *donor form-orientation* and that of *donor principle-orientation*. The former consists in using in a given MA graphic word a graph occurring in its SA cognate graphic word although this use is at odds with MA pronunciation (or morphology). For instance, ث ‹t̲› in ثاني ‹t̲any› *tani* ‘second’ is written in accordance with this rule because this graph is used to mark *t̲*, a sound absent from MA but present in the SA cognate of this word: ثانٍ ‹t̲ān^{in}› *t̲āni*^{n} ‘idem’ (cf. the MA self-oriented phonetic graphy ت ‹t› in تاني ‹tany› ‘idem’). The principle of donor principle-orientation (repetition of ‘principle’ is intended) means that a SA orthographic principle is applied to dialectal phonetic and graphic content. Thus, what is used as a model is not the form but an abstract principle. An instantiation of this principle is signalling a pseudo-prothesis by means of ا ‹a› (discussed under “ا ‹a› ~ أ ‹å›” in Sect. VI.1.). Another example are different graphies of the *hamza* sign used to mark *ʔ* depending on

the phonetic environment of this sound. For instance, the reason for using أ ‹å› in the MA graphic word مأدب ‹mådb› *mʔəddəb* 'polite' is the application of the SA orthographic principle according to which *ʔ* is marked as أ ‹å› if it is preceded by a consonant and followed by *a*, with the MA sound *ə* being treated as close enough to *a* (cf. the SA cognate, مؤدب ‹mŭdb› *muʔaddab*^{un}, in which *ʔ* is preceded by *u* and followed by *a*, hence, according to another SA orthographic rule, it is written as ؤ ‹ŭ›).

The principle of self-orientation subdivides into the phonetic principle and morphological principle, depending on the level of representation reflected in writing (as discussed in Sect. II.1.).

A phonetic graphy, i.e. one based on the phonetic principle, may be either adoptive or adaptive, depending on the genesis of a particular graph used according to this principle. The *adoptive phonetic graphy*, the dominant type in MA, means that a graph used in the donor alphabet, in the present case the SA alphabet as used in Morocco, is adopted together with its graph-sound correspondence. For example, the graphs ت ‹t› and ج ‹ǧ› mark the sounds *t* and *ž*, respectively, in MA because they do this in Moroccan SA[88]. Within the adoptive graphy, two graphic traditions have to be differentiated, reflected in the way of marking *f* and *q*. Following the Moroccan SA usage, they are commonly marked in MA as ف ‹f› and ق ‹q›, respectively, i.e. in a way *based on general donor graphy* (i.e. general Arabic graphy, the Ḥafṣ tradition, see Sect. I.1.). In addition – again, following the Moroccan SA usage – two special graphs are sometimes used in this function in the MA graphy as a regional peculiarity: ڤ ‹f̣› and ڧ ‹f$_q$›, respectively, *based on local donor graphy* (i.e. local Maghrebi graphy, the Warš tradition).

The *adaptive phonetic graphy* means that a graph is adapted for reflecting MA own phonetic purposes; i.e. the graph-sound correspondences are a matter of MA-internal conventions. Generally, adaptive phonetic graphy is used for marking sounds which have no exact counterparts in Moroccan SA. In this type of graphy, what is adapted may be a graph used in the SA graphy, i.e. *endogenous adaptation*, or a graph not occurring in this writing system, i.e. *exogenous adaptation*. An example of the former is ٳ ‹ą›, which marks in SA word-initial *ʔ* followed by *i*, but in MA it marks word-initial *i*, e.g. in إتيم ‹ątym› *itim* 'orphan'. The latter type of adaptation is exemplified by گ ‹ǵ› and ڴ ‹ĝ›, used for marking *g*.

Endogenous adaptation is subcategorized into phonetically motivated and historically motivated. *Phonetically motivated endogenous adaptation* consists in using a SA graph which, in SA, marks a sound phonetically (articulatorily and auditively) similar to a given MA sound. For instance, the MA sound *g*, with no

[88] This is not the case, for instance, in Egyptian SA, in which ج ‹ǧ› is in most cases used to mark *g*.

counterpart in Moroccan SA, is marked as ك ‹k› because this graph marks in SA *k*, a sound phonetically similar to *g*, the only difference between them being unvoicedness vs. voicedness.

Historically motivated endogenous adaptation consists in using a SA graph which marks a sound with which the MA phonetic phenomenon to be marked is diachronically related (in a way, this corresponds to the historical principle described in point 4. in Sect. I.3.). For instance, some spellers use the graph ج ‹ǧ› to mark *g* in words in which this sound is a result of dissimilaton (deaffrication) of reflexes of SA *ǧ* under the influence of a sibilant (cf. Heath 2002: 136-138; Aguadé 2008: 288). For example, *gləs* 'he sat down' (cf. SA *ǧalasa* 'idem') is often written جلس ‹ǧls›, with ج ‹ǧ›. The sounds *g* which have originated in this way are never marked as ق ‹q›, e.g. *gləs* is never spelt *قلس ‹qls›[89]. In contrast, *g* being a diachronic reflex of *q* can be marked as ق ‹q› but never as ج ‹ǧ› (e.g. *gal* 'he said' can be spelt قال ‹qal› but never *جال ‹ǧal›).

Another case of historically motivated endogenous adaptation is the use of ا ‹a›, أ ‹å› and آ ‹ã› to mark word-initial *a* in MA. In SA, in which no word begins with a vowel (e.g. Danecki 1994: 68; Corriente 2002: 35), these graphs are used to mark a word-initial glottal stop *ʔ*. Properly speaking, the *ʔalif* ا functions as support (bearer) for the *hamza* sign marking *ʔ*, e.g. أنا ‹åna› *ʔana* 'I'. However, since the *hamza* sign is frequently not written, the support has taken over the function of the sign it was to carry, i.e. ا ‹a› has been refunctionalized to mark *ʔ* so that *ʔana* 'I' is also written انا ‹ana›. In MA, which has no word-initial *ʔ* (except for borrowings from SA) and in which words do begin with a vowel, these graphs have been adapted to mark the word-initial *a*, e.g. انا ‹ana›, أنا ‹åna› and آنا ‹ãna› all represent *ana* 'I'[90]. A similar adaptation can be observed for إ ‹ạ› and ا ‹a› which word-initially mark *ʔi* in SA but *i* in MA as well as for و ‹w› which word-initially marks *w* in SA but *u* in MA.

Exogenous adaptation consists in marking a sound by means of a graph foreign to the SA alphabet, adapted from other scripts of Arabic origin (most notably Persian). To illustrate, the sound *g* in MA, apart from being marked by the SA graphs ج ‹ǧ›, ق ‹q›, ك ‹k›, is also marked by گ ‹ǵ›, a graph used, for instance, in Persian script for marking *g*, or ڭ ‹ĝ›, used, for example, in the Ottoman Turkish

[89] It also seems that ق is not used to mark *g* in loanwords. Thus, it is not likely for *gana* 'good mood' (from Spanish *gana* 'willingness') to be spelt as *قانة ‹qanħ› or *girra* 'war' (from Spanish *guerra* 'idem') as *قيرة ‹qyrħ›.

[90] One can thus propose the following refunctionalization chain of the word-initial *ʔalif*: 'support of the *hamza* sign' (SA) > 'marking *ʔ*' (SA) > 'marking a vowel' (MA).

Arabic alphabet to mark [ŋ]. These two exogenous graphs are used (together with ك ‹k›) to mark *g* irrespective of its historical development, including borrowings[91].

The dichotomous division of the spellings principles proposed for MA in the above paragraphs is represented in Figure 1.

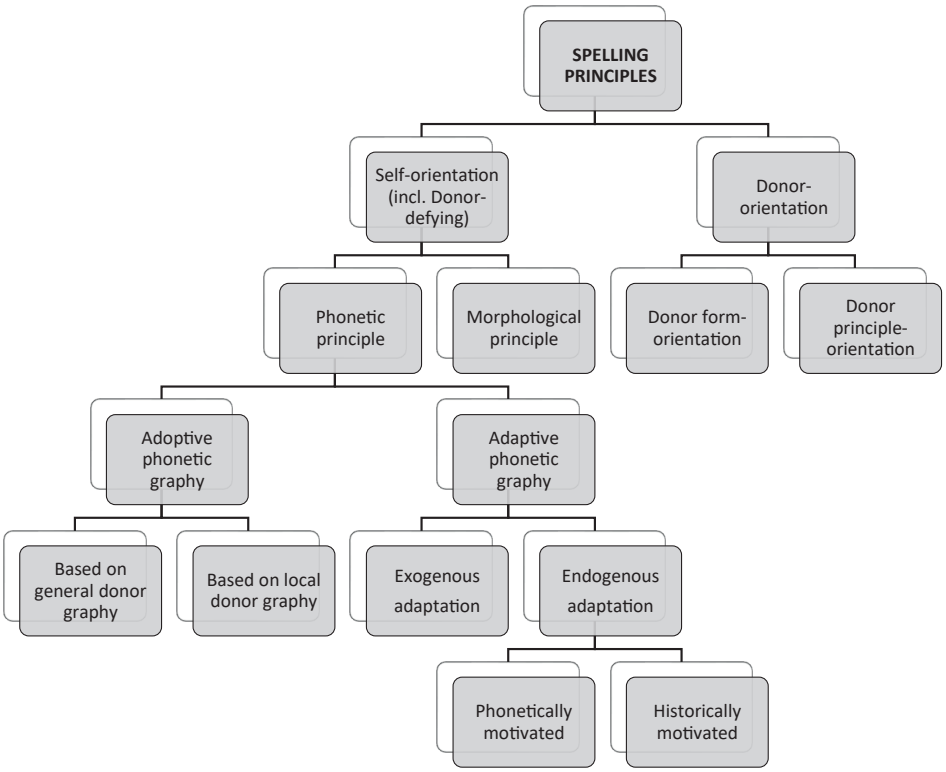


Figure 1. *Spelling principles operating in the graphy of MA*

Since some graphies cannot be explained in terms of these spelling principles, other factors responsible for them need to be identified. One of the most pervasive ones seems to be *analogy*. *Analogical graphy* is a manifestation, at the graphic level, of the linguistic phenomenon of analogy, which is based on the proportional relationship between specific linguistic forms (cf. Polański 2003b: 46). It is understood as a graphy which a speller uses in graphic word *A* because (s)he is aware

[91] Note that endogenous adaptation is responsible for the exiśtence of the phonetic ambiguity of graphs. For inśtance, ب ‹b› is adapted to mark *p* and at the same time retains its original function of marking *b*. In contraśt, graphs introduced into MA graphy *via* exogenous adaptation, e.g. پ ‹p› marking *p*, are phonetically univocal.

that it is used in graphic word *B* which is similar in some respect; however, while this graphy in *B* is justified by one of the spelling principles proposed, this is not the case for *A*. (Whether the speller is aware that, properly speaking, such graphies are based on *false* analogies is difficult to verify).

A quantitative example of analogical graphy is the use of *ʔalif al-wiqāya*, i.e. of ا ‹a› after و ‹w› marking word-final *-u*, written in forms other than plural verbs. For instance, the graphy عندوا ‹ʕndwa› (*LŠ* 54_{11}) representing the prepositional phrase *ʕənd-u* ‘he has’, lit. ‘with him’ composed of the preposition *ʕənd-* ‘with, at’ and the personal suffix *-u* ‘him’, more currently written عندو ‹ʕndw›, can be explained in the following terms: The speller is aware that the final *-u* is written as وا ‹wa› in many MA graphic words, e.g. in سمعتوا ‹smʕtwa› *səmʕtu* ‘you^{pl} heard’ and (s)he uses this graphy to mark the final *-u* in *ʕənd-u* ‘with him’. Both situations are similar in that each of these phonetic words ends in *-u* which is a suffix, but also dissimilar: *-u* in *səmʕtu* is a verbal plural suffix and its being marked as وا ‹wa› is justified by the principle of donor principle-orientation, while in *ʕənd-u* it is a personal suffix which none of the spelling principles justifies to mark in this manner.

A qualitative example of analogical graphy is the use of ذ ‹ḏ› in the graphic word الذود ‹alḏwd› (*BT* 116_3) representing *d-dud* ‘worms^{def}’. The speller uses this graphy because (s)he is aware that *d* is written ذ ‹ḏ› in many words, e.g. ذراع ‹ḏraʕ› *draʕ* ‘arm’. The situations are similar in that in each case it is the same sound and, in addition, each word has a SA cognate. However, there is a dissimilarity between them: The SA cognate of *draʕ* contains the sound *ḏ* and is written with ذ ‹ḏ› (which justifies the donor form-oriented use of this graph in MA ذراع ‹ḏraʕ›), whereas the SA cognate of MA *d-dud* ‘worms^{def}’, viz. الدود *ʔad-dūd^{u}* ‘idem’, contains *d* and is written with د ‹d›. Analogy is also responsible for this type of graphy in words that have no SA cognates at all, e.g. *druk / ḍruk* ‘now’ being written as ذروك ‹ḏrwk› (*ḌK* 37^{1}) rather than دروك ‹drwk›.

In some cases, analogy gives rise to graphies referred to as pseudo-correct in Michalski (2016: 385-386). This concerns MA words with a SA cognate for which an erroneous spelling is assumed by the speller (cf. the case of الذود ‹alḏwd› discussed above). The concept of analogical graphy helps us to understand why such graphies are used. In this study, I prefer not to use the term ‘pseudo-correct’[92] in order not to suggest that such graphies aspire to some norm but fail to reach it because of the speller’s insufficient skills. A pseudo-correct graphy may occur in

[92] Or ‘hypercorrection’, used by Rosenbaum (2004: 305-306), who described an analogous phenomenon in EA.

a text written in SA (cf. Blau 1970)[93] but not in a text written in MA, for which no orthographic norms exist.

Sometimes the choice between two graphs seems to be conditioned by the *lesser phonetic ambiguity* of one of them. This choice is, actually, a choice between two opposing principles: the principle of economy, aiming at reducing graphic elements which are not necessary for achieving communication, and the phonetic principle, requiring a maximal graph-sound correspondence[94]. In most cases, the degree of phonetic ambiguity of a graph depends on the absence or presence of elements such as dots, strokes, the *hamza* or *madda* signs, with the graph-sound correspondences being usually established in accordance with the values of the letters of the SA alphabet. The degree of phonetic ambiguity of some graphs can be described only relatively, i.e. by comparing them with another graph. In consequence, when characterizing a graph in this respect, the formulation 'graph *A* is phonetically less ambiguous than graph *B*' will be used. For instance, as far as marking the sound *a* in MA is concerned, آ ‹ã› is phonetically less ambiguous than أ ‹å›, which may also mark *u*. An even more ambiguous way of marking this sound is the graph ا ‹a›, which may also mark *u* and *i* or no sound at all. Another example is the pair ا ‹a› and ه ‹h›: the former is less ambiguous in marking the final *a* than the latter, which may also mark *u* or *h*. An aspect of this issue, which will not concern us further in this work, will be discussed in Sect. VI.2.2., after analyzing the particular graphs in more detail.

As signalled in Sect. I.2., spelling principles can overlap, i.e. the graphy of a particular graph can be interpreted as resulting from more than one underlying principle. In other words, two principles may lead to the same result. In such cases, it seems more adequate descriptively to speak of a *parallelism of spelling principles* rather than arbitrarily opt for one of them to the detriment of another. Such a parallelism can be proposed, for example, for the graphy ت ‹t› in تجي ‹tğy› *dži* (< *t-ži*) 'you come' as compared with د ‹d› used in its heterograph دجي ‹dğy›. With respect to orientation, this graphy can be interpreted as self-oriented morphological (marking the abstract morpheme *t-*) but also as donor form-oriented

---

[93] An instance of pseudo-correct graphy in Moroccan SA is the frequent case of writing the expression *mudirr*^{un} *li-r-ribḥ*^{i} 'income-generating' as مذر للربح ‹mḏr_llrbḥ›, with ذ ‹ḏ›, instead of مدر للربح ‹mdr_llrbḥ›, with د ‹d› (e.g. on a web site of the Moroccan Ministry of Education, (https://www.men.gov.ma/Ar/Pages/DetailActualite.aspx?ActuID=n/WXVdu2/hY=, accessed 30 March 2018). In this case, the speller writes ذ ‹ḏ› because (s)he is aware that certain SA words contain the sound *ḏ* and are spelt with ذ ‹ḏ›, while their MA lexical cognates contain *d*. What the speller fails to recognize is that the word in question is not one of such words (it does not even have a direct MA lexical cognate).

[94] In a different context, Coulmas (2003: 200) observes in a similar vein: "For an efficient graphic code to be developed the countervailing demands of encoding (minimizing manual work) and decoding (maximizing visual discrimination) must be taken into account".

(cf. SA تجيء ‹tğyo› *tağīʔᵘ* ‘idem’). In another instance of this phenomenon, the word-initial ا ‹a› in forms such as اجلس ‹ağls› *gləs* ‘sit down!’ does not represent any sound. One possible reason for using this graphy is donor *form*-orientation. This means that the MA graphic word results from copying the form of its SA cognate: *ʔiğlis* ‘idem’, written اجلس ‹ağls›. Another possible reason is donor *principle*-orientation, i.e. that the form is a result of applying to MA phonetics the SA orthographic *principle* which requires ا ‹a› to be written before clusters even when no prothetic vowel is pronounced; cf. for instance, the SA phonetic phrase *qāla ğlis* ‘he said: Sit down!’, in which the final *a* in *qāla* eliminates the reason for a prothetic vowel but the graph ا ‹a› marking it is retained in writing: قال اجلس ‹qal_**a**ğls›.

VI. Qualitative variational heterography and variations between graphs in Written Moroccan Arabic

In this chapter, a representative selection of qualitative variational heterograph pairs is given in which particular variations between graphs are manifested. The heterographs have been chosen from among the totality of the examined graphic words occupying the nearly 1900 printed pages included in the corpus. The primary goal of this presentation is to identify graphs that are bound by particular types of variation. In addition, graphs are also characterized with respect to the spelling principles underlying their use in order to describe the MA graphic system in a more complete way. The distinctivity approach is thus complemented by a representational approach and accompanied by an attempt at explaining the reasons behind particular graphies. Finally, in the section devoted to the graphemics of MA, the variation types identified between graphs will make it possible to propose an inventory of graphemes.

VI.1. Heterograph pairs and spelling principles

The pairs of graphs bound by variation are given and exemplified, in heterograph pairs, in a way intended to completely reflect the usage in the corpus. However, since only graphs that are in variation are characterized with respect to their functions and underlying spelling principles, the representational component of this study does not pretend to be complete. For instance, the graph ر ‹r›, which is not in any variation with any graph, is not characterized with respect to its functions (marking *r* and *ṛ*) and the spelling principles underlying its use.

For each graph affected by variation, all its variants are identified. The pairs of variants thus originating are indicated by means of a tilde, e.g. "ا ‹a› ~ أ ‹å›", and arranged in alphabetical order. Variations of each graph are analyzed with respect to positions in a graphic word: word-initially, word-medially, word-finally or in all possible positions in which the graph occurs[95]. In rare cases, there are varia-

[95] The fact that a pair of graphs is in some type of variation word-initially does not necessarily mean that it is in the same type of variation word-medially if the word-medial position is a result of graphic prefixation. For instance, word-initially, ا ‹a› can be substituted for أ ‹å› in any graphic word,

tions which manifest themselves in more than one position but for various reasons are described separately (e.g. ا ‹a› Rv ي ‹y›). Pseudo-variations are described in a similar way. In addition, the following technical rules have been observed:
1. The description of every heterograph pair and pair of variants is given under the graph which comes first in the alphabet. It is not repeated under the other graph but a cross-reference is provided instead. For instance, the asymmetrical variation between د ‹d› and ذ ‹ḏ› is described under د ‹d›. Under ذ ‹ḏ›, the reader finds a cross-reference: *See* د ‹d› Av ذ ‹ḏ›.
2. No preference is given to data from a single text or from two different texts since both situations are equally relevant. Sometimes, however, examples from a single text are given on purpose to show that variation manifests itself even in a single author's text. As far as possible, graphic words representing various parts of speech and morphological categories have been chosen as examples.
3. Some graphic words are used as examples more than once.
4. For the sake of clarity, examples with less or no vocalization have been preferred.
5. Since none of the authors uses graphies involving some kind of 'graphic sandhi', i.e. the form of a graphic word being influenced by its graphic environment, graphic words are represented in isolation.
6. Although pairs of *minimal* qualitative heterography are given priority, in exceptional cases, where pertinent examples could not be identified in the corpus, non-minimal heterograph pairs are used, provided that the additional distinctive units do not influence the graphy under consideration. The irrelevant distinctive units, whether qualitative, quantitative or linear, are transliterated in round brackets. For instance, when analyzing the word-initial graph in graphic words representing *simana* 'week', it is irrelevant whether the last graph is ا ‹a› or ة ‹ħ›, hence, the graphic words سيمانا ‹syman(a)› and صيمانة ‹ṣyman(ħ)› can be compared with each other as heterographs.
7. In sporadic cases, pairs of graphic words differing in meaning are adduced as examples provided that the meaning-differentiating elements in no way influence the graphy under discussion. These elements, referred to as *accidentals*, are usually: the definite article, conjunctions, temporal and aspectual affixes (preverbs), prepositions or number and gender affixes. They are transliterated in square brackets. For instance, the graphic words سليب ‹slyb› *slip* 'briefs' and السليپ ‹[al] slyp› 'briefs[def]' are considered comparable with each other as far as the word-final

---

e.g. أرض ‹årḍ› ~ ارض ‹arḍ› *ʔaṛḍ* 'earth', but word-medially this is not always the case since, for example, فأرض ‹fårḍ› *f-ʔaṛḍ* 'in the earth of' is not the same as فارض ‹farḍ› *faṛaḍ* 'supposing, having supposed'. This is related to the fact that while أ ‹å› word-initially is typically used to mark *a*, its function word-medially is sometimes to mark *ʔ*.

graph is concerned because any influence of the presence or absence of the definite article on its choice can be ruled out[96].
8. Expected variations together with the necessary expected forms are indicated only for cases with a high probability of their occurrence or cases that are otherwise relevant to the point. No attempt is made to indicate all expected forms or expected variations.

In some cases of variations expected to manifest themselves in all possible positions, only pairs showing variation in one or two positions have been recorded. For instance, the variation ج ‹ğ› Rv ك ‹k› has been recorded only word-initially although it is also expected to occur word-medially.

GRAPH ا ‹a›

ا ‹a› ~ أ ‹å›

Word-initially / Asymmetrical variation (ا ‹a› Av أ ‹å›)
This variation manifests itself in graphic words in which both graphs are used in one of the two following functions based, to some extent, on SA graphy:

(i) marking a word-initial vowel (or a word-initial *ʔ* followed by a vowel),

(ii) signalling initial clusters (pseudo-prothetic use) or pseudo-clusters.

The variation is asymmetrical because ا ‹a› can be substituted for أ ‹å› in all graphic words but in some graphic words أ ‹å› cannot be substituted for ا ‹a›. For instance, this is not possible in graphic words with SA cognates which are not spelt with أ ‹å›. Thus, the MA word *mtiḥan* ‘examination’ can be written امتحان ‹amṭḥan› but not أمتحان ‹åmtḥan› (cf. SA امتحان ‹amṭḥan› *ʔimtiḥān*^{un} ‘idem’).

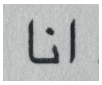
(i) An analysis of the use of ا ‹a› marking an initial vowel or a word-initial glottal stop *ʔ* followed by a vowel in MA must be preceded by a discussion of its related functions in the donor graphy. In SA, words do not begin with vowels (e.g. Danecki 1994: 68; Corriente 2002: 35). Even prothetic vowels are preceded by a glottal stop, which is elided when possible. It is a common practice in SA to mark elidible glottal stops (*hamzat al-waṣl*) by means of ا ‹a›, the bare *ʔalif*, (e.g. ادخل ‹adxl› *ʔudxul* ‘come in!’, اسم ‹asm› *ʔism*^{un} ‘name’), while non-elidible glottal

[96] An example of an element which makes two graphic words incomparable in this respect is ش ‹š› marking the negational *-š* in the pair باغية ‹baġyħ› *baġya* ‘she wants’, lit. ‘wishing^{f}’ vs. ما باغياش ‹ma_baġyaš› *ma baġya-š* ‘she does not want’, lit. ‘not wishing^{f}’. Affixed word-finally, it influences the way how the feminine affix *-a-* is marked: It requires ة ‹ħ› to be replaced by ا ‹a› because ة ‹ħ› cannot occur word-medially.

stops (*hamzat al-qaṭʕ*) are marked as أ ‹å› (the *ʔalif* with the *hamza* sign added on top) if the following vowel is *a* (e.g. أرض ‹årḍ› *ʔarḍ*^{un} 'earth') or *u* (e.g. أذن ‹åḏn› *ʔuḏun*^{un} 'ear'), and by means of إ ‹ę› (the *ʔalif* with the *hamza* sign added below) if the vowel is *i* (e.g. إثم ‹ęṯm› *ʔiṯm*^{un} 'sin')[97]. This rule is, however, often violated in SA, in two ways. On the one hand, words beginning witht a non-elidible *ʔ* are written with ا ‹a› (e.g. ارض ‹arḍ› *ʔarḍ*^{un} 'earth') by "some less conscientious typographers" (Beeston 1970: 26). On the other hand, the elidible *ʔ* is (rather rarely, mostly for didactic purposes and utterance-initially) written as أ ‹å› or إ ‹ę›, depending on the following vowel (e.g. أدخل ‹ådxl› *ʔudxul* 'come in!', إسم ‹ęsm› *ʔism*^{un} 'name')[98].

This inconstinency has been borrowed into the MA graphy, albeit with different functions: Since MA words do not begin with *ʔ*, except for borrowings from SA, the primary function of ا ‹a› and أ ‹å› is to mark a word-initial *a*, as shown in (1)-(3). Both MA graphies are based on the adaptive phonetic principle, with the adaptation being endogenous and historically motivated. In (1), which shows a graphic word with a SA cognate, both graphies are coincident with the donor graphy.

(1)

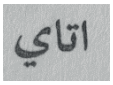


‹**a**na› (*TN* 42^{5})

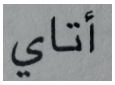
‹**å**na› (*MX* 9$_{2}$)

ana 'I' (cf. SA أنا ‹åna› and انا ‹ana› *ʔana* 'idem')

(2)



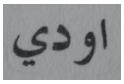
‹**a**tay› (*ḤB* 14$_{1}$)



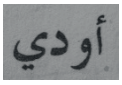
‹**å**tay› (*TN* 26^{6})

atay 'tea'

(3)



‹**a**wdy› (*ḤB* 47^{12})



‹**å**wdy› (*ḤB* 15$_{6}$)

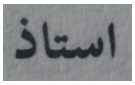
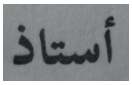
a-wədd-i 'pal!, my dear!' (*a-* is a vocative particle)

---

[97] Such a rule is given indirectly in, for instance, al-ʔAsmar (1988: 130) and Qabbiš (1984: 26). However, many Arab authors of normative spelling books fail to explicitly mention any rules underlying the choice between ا ‹a› or أ ‹å›/إ ‹ę› in the word-initial position, although they apply them in their examples (e.g. Hārūn 1993: 7-11; al-Ġalāyīnī 2002 [1912]: 269-272).

[98] Al-ʔAsmar (1988: 130) advocates such spellings as a means of facilitating reading. Badawi *et al.* (2004: 12) observe that this "innovation (…) has largely been sanctioned by the language academies".

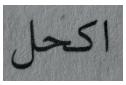
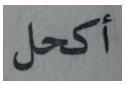
Marking word-initial *ʔ* followed by *a* or *u* is a secondary function of these graphs in MA, observable only in borrowings from SA. Such graphies are adoptive phonetic and coincident with the donor graphy. Example:

(4)  

‹**a**staḏ› (*Mu.* 30[8]) ‹**å**staḏ› (*Mu.* 30[5])

ʔŭstad ‘professor’ (cf. SA أستاذ ‹åstaḏ› and استاذ ‹astaḏ› *ʔustāḏ*[un] ‘idem’)

(ii) Before analyzing the second function of ا ‹a› in MA, i.e. the one of signalling initial clusters, including geminates, and pseudo-clusters, it is necessary to outline the issue of the syllabic structure of the onsets of MA words. According to most descriptions, a word can begin with a two-consonant cluster, i.e. with no prothetic vowel. Some authors state it explicitly (e.g. Kjamilev 1968: 33 and Durand 2004: 61-62, who even posits three-consonant initial clusters), while others suggest it through their transcription (e.g. Harrell 1962; Youssi 1992; Aguadé & Benyahia 2005)[99]. Despite this, the graph ا ‹a› is very often written at the beginning of graphic words representing such words (cf. Aguadé 2006: 257, pt. 4.1[100]). Rarely, أ ‹å› is used instead (آ ‹ã› has not been recorded in this function). Some graphies of this type can be considered to be donor form-oriented, e.g. (5) and (6).

(5)  

‹**a**kḥl› (*BB* 91[1]) ‹**å**kḥl› (*ḤB* 33[1])

kḥəl ‘black’ (cf. SA أكحل ‹åkḥl› and اكحل ‹akḥl› *ʔakḥal*[u] ‘black’, said of eyes)

(6)  

‹**a**nt› (*ḌK* 13[5]) ‹**å**nt› (*BB* 44[10])

nta ‘you’ (cf. SA أنت ‹ånt› and انت ‹ant› *ʔanta* ‘idem’)

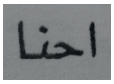
---

[99] The alternative view, according to which a prothetic *shwa* appears in such situations, is suggested by Abdel-Massih’s (1974: 9-10) transcription: [ᵊl-wᵊld] ‘boy[def]’, [ᵊš-šᵊṛžᵊm] ‘window[def]’.

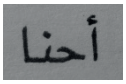
[100] Hoogland (2013a: 71) writes that in two of his sources, the *ʔalif* is used “instead of a *ya* before verbs, both in the past and present tense”. However, he fails to recognize two different uses of this graph: one is the marking of the present tense prefix in verbs and the other one is signalling a consonant cluster (in the latter case, this graph is not used instead of a *ya*, as revealed by the examples he adduces).

However, many MA graphic words with word-initial ا ‹a› or أ ‹å› have no SA cognates written in this way, for instance:

(7)



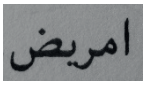
‹**a**ḥna› (*Bə.* 76[1])



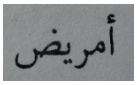
‹**å**ḥna› (*Bə.* 76_{11})

ḥna ‘we’ (cf. SA نحن ‹nḥn› *naḥnu* ‘idem’)

(8)



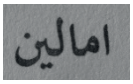
‹**a**mryḍ› (*XM* 149[8])



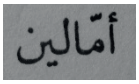
‹**å**mryḍ› (*Ṣa.* 229_{7})

mṛiḍ ‘ill’ (cf. SA مريض ‹mryḍ› *marīḍ*[un] ‘idem’)

(9)



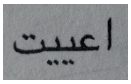
‹**a**malyn› (*ḤB* 6[9])



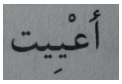
‹**å**mʿalyn› (*ḤD* 23_{1})

ṃṃʷalin, *mwalin* ‘owners’

(10)



‹**a**ʕyyt› (*ṮN* 44[6])



‹**å**ᵃʕʿyⁱyt› (*MX* 82_{1})

ʕyit ‘I got tired’ (cf. SA عييت ‹ʕyyt› *ʕayiytu* ‘I was unable’)

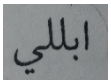
These graphies might be understood as reflecting the feeling of some MA spellers that a prothetic vowel is pronounced. However, a more probable explanation seems to be the principle of donor principle-orientation: According to the SA principle applied here, an *ʔalif* must be written word-initially in SA words beginning with a two-consonant cluster, even when this vowel is not pronounced due to elision. Following this rule, the pseudo-prothetic ا ‹a› can be written in MA even where a consonant cluster is preceded by a word ending in a vowel, which would cause the elision of any potential prothetic vowel. For instance, *ši flus* ‘some money’, usually spelt شي فلوس ‹šy_flws›, can also be written as شي افلوس ‹šy_**a**flws› (*XM* 130[3])[101].

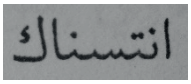
The use of أ ‹å› in this function is an instance of analogical graphy: The speller assumes that since ا ‹a› can be substituted for by أ ‹å› in some contexts (e.g. انا

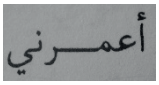
[101] Al-Fāsī (1986), explains his using ا ‹a› word-initially in MA graphic words such as اجلس ‹ağls› *gləs* ‘sit down!’, where it does not represent any sound, by comparing it to the SA *ʔalif al-wiqāya* in verbal forms: ‘it is like the *ʔalif* which we put at the end of the word دخلوا ‹dxlwa› *daxalū* ‘they entered’’ (1986: 24). Unfortunately, he fails to indicate its function. He may simply mean that what both uses of ا ‹a› have in common is the absence of a phonetic function.

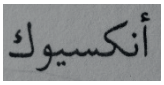
‹ana› and أنا ‹åna› for *ana* ‘I’), this may also occur in graphic words representing, for instance, *ḥna* ‘we’, *mṛiḍ* ‘ill’, et cetera, although the phonetic circumstances are different.

The second cluster-related function of the initial ا ‹a› and أ ‹å› is the signalling of pseudo-clusters. The term ‘pseudo-cluster’ is used here to refer to a sequence of three word-initial consonants which is split by a reduced vowel between the first and second consonant (*shwa* in Benhallam 1980, 1990, as cited in Boudlal 2011, and *ə* or, in some environments, *ŭ* in Harrell 1962 and Aguadé & Benyahia 2005), resulting in the structures *CəCC-* or *CŭCC-*[102]. Speech pace seems to be of importance in this respect: In careful speech, reduced vowels can be heard, while rapid speech may eliminate them. At any rate, ا ‹a› and أ ‹å› are used in graphic words representing this type of words in one source only, the author of which most probably wanted to signal what he considered to be initial clusters. Since no heterograph pairs could be established, the following examples are given without heterographs: (11) and (12) show ا ‹a› used in this function, while in (13) and (14) the use of أ ‹å› is exemplified[103]:

(11) 
‹**a**blly› (*Ṣa.* 245[5])
bəlli ‘that’ (conjunction)

(12) 
‹**a**ntsnak› (*Ṣa.* 222[3])
nətsənna-k ‘I wait for you’

(13) 
‹**å**ʕmrny› (*Ṣa.* 215[8])
ʕəmməṛ-ni ‘I never’

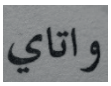
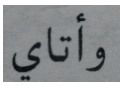
(14) 
‹**å**nksywk› (*Ṣa.* 212$_{10}$)
nəksiw-ək ‘we dress you’

The use of ا ‹a› in this function can thus be explained as a result of the principle of donor principle-orientation. Its variant أ ‹å› is used by way of analogical graphy, as was the case with clusters discussed above.

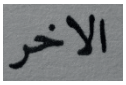
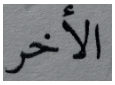
[102] Durand (2004: 58, 61-64) holds a different view, arguing that three word-initial consonants are pronounceable without a *shwa*. The latter, in his opinion, is given in “traditional notation” for merely didactic purposes. However, he admits that some “additional interruptions” (“ulteriori cesure”) which facilitate pronunciation can be heard, without constituting fully-fledged syllables (he terms them ‘pseudosyllables’, p. 63).

[103] The word in (13) is utterance-initial while the remaining ones are not but it is difficult to tell if this is of any relevance.

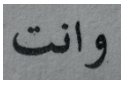
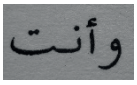
Word-medially / Restricted variation (ا ‹a› Rv أ ‹å›)
In graphically prefixed words, the graphs ا ‹a› and أ ‹å› have the same function as word-initially (q.v.): marking a vowel (rarely, preceded by *ʔ*) or signalling a pseudo-prothesis. If there is no SA cognate, both graphies are adaptive phonetic, with the adaptation being endogenous and historically motivated. Example:

(15)  
‹w**a**tay› (*Ḥ̣B* 6^{5}) ‹w**å**tay› (*MX* 49_{4})
w-atay ‘and tea’

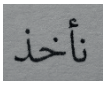
If the word has a SA cognate spelt with آ ‹ã›, the principle underlying أ ‹å› seems to be donor-defying. For example:

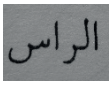
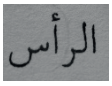
(16)  
‹al**a**xr› (*Ḥ̣D* 28_{5}) ‹al**å**xr› (*LŠ* 96_{3})
l-axŭ̱r ‘other^{def}’ (cf. SA الآخر ‹alãxr› or الاخر ‹alaxr› *ʔal-ʔāxar^{u}* ‘idem’)

In some graphically prefixed words, both graphs can be interpreted as donor form-oriented or as remnants of word-initial donor form-oriented graphy. For instance:

(17)  
‹w**a**nt› (*Ḥ̣B* 48^{2}) ‹w**å**nt› (*Mğ.* 28_{5})
u-nta ‘and you’
(cf. SA وانت ‹want› and وأنت ‹wånt› *wa-ʔanta* ‘idem’ and MA انت ‹ant› and أنت ‹ånt› *nta* ‘you’)

In words that are not graphically prefixed, ا ‹a› used to mark *a* is an instance of adoptive phonetic graphy based on general donor graphy, whereas أ ‹å›, reflecting *ʔ* of the SA cognates, is used following the principle of donor form-orientation. Examples:

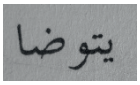
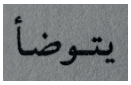
(18)  
‹n**a**x̱ḏ› (*BT* 116_{3}) ‹n**å**x̱ḏ› (*Ṣa.* 238^{6})
naxŭd ‘I take’ (cf. SA نأخذ ‹nåx̱ḏ› *naʔxuḏ^{u}* ‘we take’)

(19)  

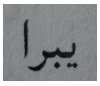
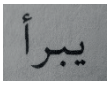
‹alr**a**s› (*Ṣa.* 238[9]) ‹alr**å**s› (*Ṣa.* 238[9])
ṛ-ṛaṣ ‘head[def]’ (cf. SA الرأس ‹alrås› *ʔar-raʔsᵘ* ‘idem’)

This variation is restricted because, despite a large number of contexts in which ا ‹a› can be substituted for أ ‹å› word-medially, this cannot happen in graphic words representing some phonetic words borrowed from SA with a word-medial glottal stop. A quite consistent practice is to mark this sound as أ ‹å›, an adoptive phonetic graphy based on general donor graphy. Examples: نأكّد ‹nåkːd› *nʔəkkəd* ‘I confirm’ (*TN* 43$_{10}$), فأرض ‹fårḍ› *f-ʔəṛḍ* ‘in the earth of’ (*Mu.* 29[4]), the latter instance being a graphically prefixed word. No heterographs with ا ‹a› have been recorded for them. If found, such graphic words would most probably have different meanings, cf. ناكد ‹nakd› *nakəd* ‘sad’ or فارض ‹farḍ› *faṛəḍ* ‘supposing, imagining’[104].

Word-finally / Asymmetrical variation (ا ‹a› Rv أ ‹å›)
This variation manifests itself in verbs that have SA cognates with the word-final أ ‹å› marking a glottal stop. The graphy ا ‹a› is adoptive phonetic, based on general donor graphy, while the use of أ ‹å› is donor form-oriented. Examples:

(20)  

‹ytwḍ**a**› (*Fʕ* 18[11]) ‹ytwḍ**å**› (*Ḍa.* 97[7])
yĭtwŭḍḍa ‘he performs his ablutions’ (cf. SA يتوضّأ ‹ytwḍːå› *yatawaḍḍaʔᵘ* ‘idem’)

(21)  

‹ybr**a**› (*MX* 112$_{12}$) ‹ybr**å**› (*BB* 70[8])
yəḅṛa ‘he gets cured’ (cf. SA يبرأ ‹ybrå› *yabraʔᵘ* ‘idem’)

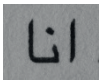
[104] This reasoning is not invalidated by the fact that for the graphic word مسألة ‹msålħ›, used normally to represent *məsʔala* ‘issue, matter’, a heterograph with ا ‹a› has been recorded: مسالة ‹msalħ› (*QQ* 57[8]) – a *hapax legomenon* being doubtless a result of careless graphy rather than a deliberate choice.

ا ‹a› ~ آ ‹ã›

Word-initially / Asymmetrical variation (ا ‹a› Av آ ‹ã›)

This variation manifests itself in graphic words in which both ا ‹a› and آ ‹ã› mark word-initial *a* (see discussion under ا ‹a› Av أ ‹å›). Seeing that آ ‹ã› cannot be substituted for ا ‹a› in some graphic words, e.g. it is never used pseudo-prothetically (e.g. no *آحنا ‹ãḥna› for *ḥna* 'we' or *آبللي ‹ãblly› for *bəlli* 'that' (conjunction) have been recorded), the variation is asymmetrical. Both graphs used in the function of marking word-initial *a* are adaptive phonetic graphies (in SA, they mark the word-initial *ʔ*), with the adaptation being endogenous and historically motivated. If the word has a SA cognate written with أ ‹å› or ا ‹a›, as in (22), آ ‹ã› may be additionally considered donor-defying, while ا ‹a› coincident with SA graphy. Examples:

(22)

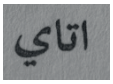


‹**a**na› (*TN* 42[5])

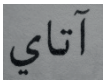
‹**ã**na› (*Bə.* 45[3])

ana 'I' (cf. SA أنا ‹åna› and انا ‹ana› *ʔana* 'idem')

(23)



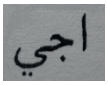
‹**a**tay› (*ḤB* 14$_1$)



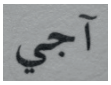
‹**ã**tay› (*MX* 81$_{12}$)

atay 'tea'

(24)



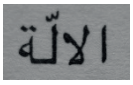
‹**a**ğy› (*ḌK* 35$_3$)



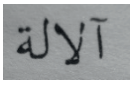
‹**ã**ğy› (*ḤB* 9$_3$)

ažı 'come!; hey!'

(25)



‹**a**lal·ħ› (*TN* 55$_2$)

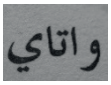
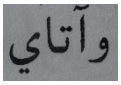


‹**ã**lalħ› (*DB* 251[1])

a-lalla 'madam!' (*a-* is a vocative particle)

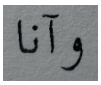
Word-medially / Restricted variation (ا ‹a› Rv آ ‹ã›)

Unquestionable cases of mutual substitutability between ا ‹a› and آ ‹ã› occur only in graphic words in which these graphs become word-medial as a result of graphic prefixation: in (26)-(27) the prefix is a conjunction, in (28) the definite article. Like in the case of their word-initial variation (q.v.), both graphies are adaptive phonetic, with the adaptation being endogenous and historically motivated. The following example shows a word with no SA cognate:

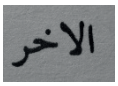
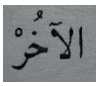
(26)  

‹w**a**tay› (*Ḥ*B 6$_5$) ‹w**ã**tay› (*MX* 81$_{15}$)
w-atay ‘and tea’

If there is a SA cognate, some graphies are also coincident with the donor graphy, e.g. ا ‹a› in (27) and both graphies in (28), while others are donor form-defying, e.g. آ ‹ã› in (27).

(27)  

‹w**a**na› (*T*M 11$_{12}$) ‹w**ã**na› (*DB* 142$_3$)
w-ana ‘and me’ (cf. SA وأنا ‹wånа› and وانا ‹wana› *wa-ʔana* ‘idem’)

(28)  

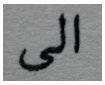
‹al**a**xr› (*Ḥ*D 28$_5$) ‹al**ã**x^{u}r^{ə}› (*Mm.* 111$_2$)
l-axŭṛ ‘other^{def}’ (cf. SA الآخر ‹alãxr› and الاخر ‹alaxr› *ʔal-ʔāxar^{u}* ‘idem’)

This variation is classed here as restricted, not asymmetrical, because in contexts other than graphically prefixed words, it does not occur or its status is dubious, for two reasons: The first one is that it cannot be ruled out that each of the two graphies reflects a different pronunciation. For instance, كيامن ‹[k]yamn› ‘he believes’ (*ṢX* 97^7) may represent *yamən* (a form given by Aguadé & Benyahia 2005: 29 and Sánchez 2014: 131), while يآمن ‹yãmn› (*Mi.* 153^1) may represent *yʔamən* (a form given by Harrell 1966: 2). The second difficulty is the scarcity of data: Graphic words with a word-medial آ ‹ã› are rare and their heterographs with ا ‹a› even rarer (recorded only twice, in *ṢX*). No heterographs with ا ‹a› for such words as كنآمر ‹knãmr› *ka-nʔaməṛ* ‘I order, I command’ (*Mn.* 45^2) or كنآذن *ka-nʔadən* ‹knãd̲n› *ka-nʔadən* ‘I permit’ (*Bə.* 83$_9$) have been recorded[105]. For all these reasons, it is more adequate to consider that آ ‹ã› cannot be substituted for by ا ‹a› in some graphic words.

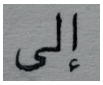
[105] For the frequent word *l-Qŭṛʔan* ‘the Quran’, a borrowing from SA consistently pronounced with a glottal stop and written القرآن ‹alqrãn›, one heterograph القران ‹alqran› (*Mm.* 149^8) has been identified. However, this *hapax legomenon* appears to be a misprint or careless graphy rather than a deliberate graphic decision.

Word-initially / Asymmetrical variation (ا ‹a› Av ع ‹ą›)
This variation manifests itself in graphic words in which ا ‹a› and ع ‹ą› mark word-initial *i* or *ǐ* (cf. discussion under "ا ‹a› Av أ ‹å›") as well as *y* (discussed below). Used in the function of marking *i* or *ǐ*, they are adaptive phonetic graphies (in SA, they mark the initial *ʔ*), with the adaptation being endogenous and historically motivated. Such graphies have been recorded in graphic words with no SA cognates, as shown in (29)-(30).

(29)

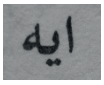


‹**a**lá› (*Ḍa.* $_{6}$44)
ila 'if'

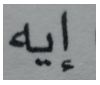


‹**ą**lá› (*Ṯ N* 8[13])

(30)



‹**a**yh› (*Ḥ B* 13$_{2}$)
ǐyyəh 'yes'



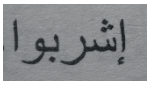
‹**ą**yh› (*Bə.* 86[4])

Both graphs are also used to mark the 3rd person non-feminine present tense, although neither has this function frequently (ا ‹a› is rather idiosyncratic of *Ḍa.*, while ع ‹ą› is used, quite consistently, in *Ṛḥ.*). Two principal manners of spelling this prefix have been recorded in the corpus: ع ‹ą›, and its less frequent variant ا ‹a› (exceptionally also أ ‹å›), is one of them. The second is ي ‹y›[106]. Before the two graphies can be analyzed, a discussion of the pronunciation of this prefix is necessary. According to Kjamilev (1968: 40), it is *y-* if the stem begins with a vowel, e.g. *yəktəb* 'he writes' (the stem being *-əktəb*), and *i-* if the stem begins with a consonant, e.g. *išuf* 'he sees'. The issue is viewed in a similar way by Aguadé & Benyahia (2005: 9), while Harrell (1962: 47-48) presents a more complex description: The prefix is usually *i-* (e.g. *iddi* 'he takes'), but before stems beginning with *w*, *h*, *ʕ* or *ḥ* followed by a consonant it is usually *yə-* (e.g. *yəwṣəl* 'he arrives', *yəhḍəṛ* 'he talks', *yəḥməl* 'he carries', *yəʕṭiw* 'they give'). However, he adds, "[i]n slower speech, especially for words pronounced in isolation, the prefix may take the form *yə-* before any consonant cluster (…), e.g. *yədxol* 'he enters'". Thus, what Kjamilev and Aguadé & Benyahia formulate as a general rule, Harrell gives only for slower speech. According to Sánchez (2014: 117), the prefix is *y(ə)-* (in the Marrakesh variety).

[106] A third one is their combination: ايـ ‹ay›, إيـ ‹ąy›, or أيـ ‹åy›, not discussed here seeing that it involves quantitative variation.

Hence, when determining the pronunciation of this prefix, the structure of the stem and the pace of speech should be taken into consideration (with the role of the latter being impossible to verify in the case of written texts). At any rate, two major graphies, ا ‹a›/ٱ ‹ą› and ي ‹y› are in use. Each of them is better suited to mark a particular pronunciation, *i*- and *y*-, respectively, by virtue of the graph-sound correspondences obtaining in SA. For the needs of this description, an assumption is made that these are their *typical* functions as far as representing the prefix in question in writing is concerned, although usage in some sources does not confirm this – admittedly idealistic – rule (cf. also "ا ‹a› Rv ي ‹y›")[107].

Let us consider the two following examples of variation between ا ‹a› and ٱ ‹ą›. In order to exclude the possibility that the graphy of the prefix is conditioned by the consonantal or vocalic character of a preceding sound, every heterograph pair comprises words which occurred in the same phonetic environment: Both verbs in (31) are preceded by the semivowel *w*, and both in (32) are preceded by a vowel.

(31)  

‹**a**šʿrʲbʿw› (*Ḍa.* 47$_4$) ‹**ą**šrbw(a)› (*Ṛḥ.* 172^6)

išəṛbu 'they drink' (cf. SA يشربوا ‹yšrbwa› *yašrabū* 'that they drink')

(32)  

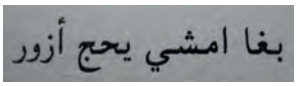
‹**a**qtʕ› (*Ḍa.* 76^8) ‹**ą**qʿtʕ› (*Ḍa.* 76^5)

yəqtəʕ 'he cuts' (cf. SA يقطع ‹yqtʕ› *yaqṭaʕᵘ* 'idem')

In the light of the assumption made above, both ا ‹a› and ٱ ‹ą› in (31) marking *i* can be considered adaptive phonetic graphies, with the adaptation being endogenous and historically motivated. As for (32), both graphs could be interpreted as donor-defying, but it seems more convincing to regard them as analogical graphies: The

---

[107] The graphy of this prefix does seem to depend on the syllabic structure of the stem in, for instance, *Ṛḥ*. In *Ḍ*, by contrast, a factor which may influence it, at least to a certain extent, is the phonetic environment, more specifically: the sound preceding the prefix. Inconsistency, however, is strong there too, as the following sequence of three verbs with various graphies of the prefix shows:



‹bġa_**a**mšy_**y**ḥğ_**ą**zwr› (*Ḍa.* 97^9)

bġa yĭmši iḥəžž iẓuṛ

'he wants to go on a pilgrimage to visit…', lit. 'he wanted he goes he makes a pilgrimage he visits'

speller is aware that this MA verbal prefix is marked as ا ‹a› and ِ ‹ą› in some cases and (s)he uses this graphy in a verb with the prefix being pronounced in a way which does not justify this. The analogical graphy may be additionaly induced by the relatively small auditive difference between the two pronunciations.

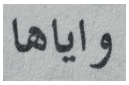
Word-medially / Asymmetrical variation (ا ‹a› Av ِ ‹ą›)
This variation manifests itself only in graphic words in which the word-medial position of these graphs results from graphic prefixation. Otherwise, the graph ِ ‹ą› never occurs word-medially in MA. Examples:

(33)

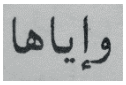
‹w**a**lá› (*Mn.* 42$_{11}$)
w-ila ‘and if’

‹w**ą**lá› (*Mn.* 18$_{4}$)

(34)



‹w**a**yaha› (*ḤB* 18$_{6}$)
w-ǐyya-ha ‘with her’ (cf. SA وإياها ‹wąyaha› and واياها ‹wayaha› *wa-ʔiyyā-hā* ‘idem’)

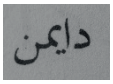


‹w**ą**yaha› (*MX* 64^{12})

If the graphs ا ‹a› and ِ ‹ą› are interpreted as marking the sounds *i* or *ǐ*, they are adaptive phonetic graphies (in SA, they mark *-ʔi-*), with the adaptation being endogenous and historically motivated. In (34), for which a SA cognate exists, both graphies can be interpreted as donor form-oriented.

ا ‹a› ~ ن ‹n›

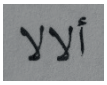
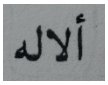
Word-finally / Restricted variation (ا ‹a› Rv ن ‹n›)
This variation manifests itself, very marginally, in graphic words which represent set expressions ending in *-ən* and borrowed from SA. In SA, the ending is *-an*; it originally signals the accusative and is written in most cases by means of ً ‹a^{an}›, the so-called *tanwīn ʔalif* (ا ‹a› with a double *fatḥa*), or simply ا ‹a›. The latter graphy in *A* in (35) is thus donor form-oriented, while ن ‹n› in *B* is adoptive phonetic, based on general donor graphy. It is the only occurrence of this type of graphy recorded in the corpus.

(35)  

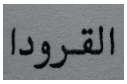
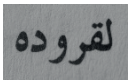
‹daym**a**› (*Bə.* 71_2) ‹daym**n**› (*Fʕ* 58_{11})
daymən ‘always’ (cf. SA دائماً ‹daỷma[an]› or دائما ‹daỷma› *dāʔiman* ‘idem’)

ا ‹a› ~ ه ‹h›

Word-finally / Restricted variation (ا ‹a› Rv ه ‹h›)
The graph ه ‹h› is in variation with ا ‹a› by virtue of its being a graphically reduced variant of ة ‹ħ› (see “ه ‹h› Rv ة ‹ħ›” for details). Hence, this quite frequent variation manifests itself in the same contexts as ا ‹a› Rv ة ‹ħ›, i.e. in some words ending in *a*, including nouns, adjectives, adverbs, but not verbs (which are written only with a final ا ‹a›, أ ‹å› or ى ‹á›; see, however, footnote 110). To some extent, it mirrors the usage in SA, where some borrowings, in particular proper names, have variational heterographic spellings, e.g. أوروبا ‹åwrwba› ~ أوروبه ‹åwrwbh› for *ʔurubba* ‘Europe’[108] (the spelling with ا ‹a› being considerably more frequent). Some of the MA graphic words have SA cognates, which makes their graphy relativizable to the donor form; for instance ا ‹a› in (40) can be considered donor-defying, while ه ‹h› coincident with the donor graphy. However, it seems more adequate to describe ه ‹h› (and ة ‹ħ›, below) as having been refunctionalized within SA graphy to mark *a* (cf. Sect. I.3., point 4.e. and discussion under “ه ‹h› Rv ة ‹ħ›”) and, in consequence, to characterize both graphies ا ‹a› and ه ‹h› as adoptive phonetic, based on general donor graphy. Examples:

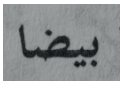
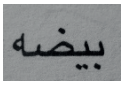
(36)  

‹ålal**a**› (*Fʕ* 8_2) ‹ålal**h**› (*XM* 130^2)
a-lalla ‘madam!’

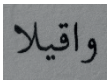
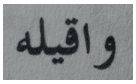
(37)  

‹(a)lqrwd**a**› (*QQ* 38_6) ‹lqrwd**h**› (*Mu.* 42_{11})
lə-qṛuḍa ‘monkeys[def]’

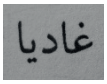
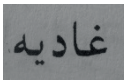
[108] The final vowel *a* in this word is not long but neutral with respect to quantity, with the graph ا ‹a› being used as a *mater lectionis* to indicate its quality.

(38)  

‹byḍ**a**› (*ḤB* 22_{6}) ‹byḍ**h**› (*ṮN* 56^{9})
biḍa ‘white^{f}’ (cf. SA بيضاء ‹byḍao› *bayḍāʔ*^{u} ‘idem’)

(39)  

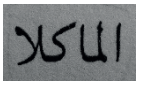
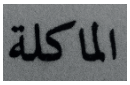
‹waqyl**a**›(*Ṣa.* 231^{11}) ‹waqyl**h**› (*Mu.* 37_{15})
waqila ‘perhaps’

(40)  

‹ġady**a**› (*Ṣa.* 227^{10}) ‹ġady**h**› (*MX* 151^{1})
ġadya ‘going to^{f}’ (future marker)
(cf. SA غادية ‹ġadyħ› and غاديه ‹ġadyh› *ġādiya*^{tun} ‘becoming^{f}’)

ا ‹a› ~ ة ‹ħ›

Word-finally / Restricted variation (ا ‹a› Rv ة ‹ħ›)
This variation mirrors the preceding one because ة ‹ħ› is a variant of ه ‹h› in the function of marking the final *a*. The analysis given under “ا ‹a› Rv ه ‹h›” applies here as well, with the corresponding variational heterograph pair in SA being أوروبا ‹åwrwba› ~ أوروبة ‹åwrwbħ› for *ʔurubba* ‘Europe’[109] (the spelling with ا ‹a› being considerably more frequent). Like in the case of ا ‹a› Rv ه ‹h›, this very frequent variation does not manifest itself in verbs or pronouns, which are not spelt with a word-final ة ‹ħ›[110]. Following the analysis given under “ا ‹a› Rv ه ‹h›”, both graphies ا ‹a› and ة ‹ħ› used in MA are interpreted as being adoptive phonetic, based on general donor graphy, with ة ‹ħ› having been refunctionalized within SA graphy to mark *a*. Examples:

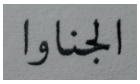
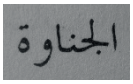
(41)  

‹almakl**a**› (*Mm.* 147^{8}) ‹almakl**ħ**› (*ḤB* 4_{10})
l-makla ‘food, eating^{def}’

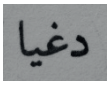
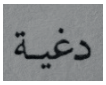
---

[109] See preceding footnote.

[110] One exception, however, has been recorded: اتّسالة ‹atˑsalħ› *tətsala* ‘it^{f} ends’ (*Ṛḥ.* 210_{5}), instead of, for instance, اتّسالا ‹atˑsala› or اتّسالى ‹atˑsalá›. A similar example can be adduced for EA: تبقة ‹tbqħ› *tibʔa* ‘you are’ (*ʕAyz.* 9^{2}), instead of the usual graphy تبقى ‹tbqá›. Such occurrences are yet another argument for the view that ة ‹ħ› and ه ‹h› have been refunctionalized to mark the final *a*.

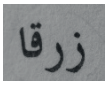
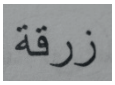
(42)  

‹alğnaw**a**› (*Mm.* 128[5]) ‹alğnaw**ħ**› (*Fʕ* 16[3])
lə-žnawa ‘knives[def]’

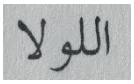
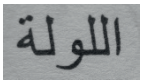
(43)  

‹dġy**a**› (*ḤD* 15[5]) ‹dġy**ħ**› (*Mn.* 78$_4$)
dəġya ‘quickly’

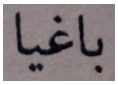
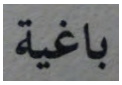
In some graphic words with SA cognates, exemplified below, the phonetic use of ا ‹a› or ة ‹ħ› can also be donor-defying (both graphies in 45, ا ‹a› in 46-48) or coincident with the donor graphy (ة ‹ħ› in 46-48).

(44)  

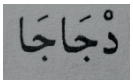
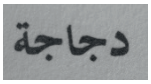
‹zrq**a**› (*ḤB* 12$_8$) ‹zrq**ħ**› (*Ṛḥ.* 163[4])
ẓəṛqa ‘blue[f]’ (cf. SA زرقاء ‹zrqao› *zarqāʔᵘ* ‘idem’)

(45)  

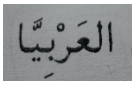
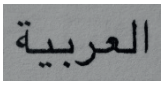
‹allwl**a**› (*Fʕ* 4$_6$) ‹allwl**ħ**› (*Ṛḥ.* 148$_4$)
l-lŭwwla ‘first[f def]’ (cf. SA الأولى ‹alåwlá› *ʔal-ʔūlā* ‘idem’)

(46)  

‹baġy**a**› (*ṢX* 96[2]) ‹baġy**ħ**› (*ḤB* 18$_7$)
baġya ‘she wants’, lit. ‘wishing[f]’ (cf. SA باغية ‹baġya› *bāġiyaᵗᵘⁿ* ‘desiring[f]’)

(47)  

‹dğ^{a}ağ^{a}**a**› (*Mm.* 111[2]) ‹dğağ**ħ**› (*Mu.* 5[1])
džaža ‘hen’ (cf. SA دجاجة ‹dğağħ› *daǧāǧaᵗᵘⁿ* ‘idem’)

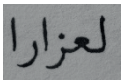
(48)  

‹alʕ^{a}r̆b^{i}y$^{:a}$**a**› (*Mm.* 121[7]) ‹alʕrby**ħ**› (*ṮN* 13[13])
l-ʕəṛbĭyya ‘Arabic language[def]’ (cf. SA العربية ‹alʕrbyħ› *ʔal-ʕarabiyyaᵗᵘ* ‘idem’)

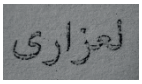
Word-finally / Restricted variation (ا ‹a› Rv ى ‹á›)

This very frequent variation manifests itself in graphic words belonging to various semantic categories, including verbs, in which both graphs mark final *a*. It mirrors the same variation in SA, which, by contrast, manifests itself in a very limited number of words (e.g. borrowings such as *mūsīqā* 'music', written موسيقا ‹mwsyqa› or, much more frequently, موسيقى ‹mwsyqá›). In MA, both graphies, ا ‹a› and ى ‹á›, are adoptive phonetic, based on general donor graphy. In (49)-(54), graphic words with no SA cognates are exemplified.

(49)

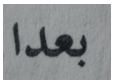


‹lʕzar**a**› (*ṮM* 31_{7})
lə-ʕzara 'bachelors^{def}'

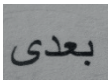


‹lʕzar**á**› (*BT* 15_{1})

(50)



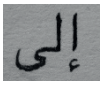
‹bʕd**a**› (*ḤB* 7^{8})
bəʕda 'already; first'



‹bʕd**á**› (*Ṛḥ.* 157_{3})

(51)

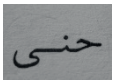
‹ḷ**a**› (*ḤB* 4^{9})
ila 'if'



‹ḷ**á**› (*ṮN* 8^{13})

(52)

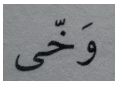
‹ḥn**a**› (*ṢX* 95_{4})
ḥna 'we'



‹ḥn**á**› (*ṮM* 65^{9})

(53)

‹wx**a**› (*Ṣa.* 209^{8})
waxxa 'although'



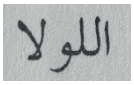
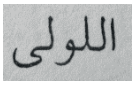
‹w^{a}x$^{\cdot}$**á**› (*BB* 107_{6})

(54)

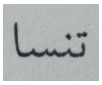
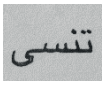
‹år**a**› (*MX* 13_{7})
aṛa 'give!' (a verb-like particle)

‹år**á**› (*MX* 46_{13})

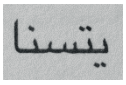
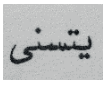
The graphy of words with SA cognates can also be relativized to the donor graphy. In (55)-(59), the adoptive phonetic graphy ا ‹a› is donor form-defying, while ى ‹á› is coincident with the donor graphy.

(55)  

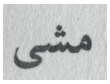
‹allwl**a**› (*Fʕ* 4_6) ‹allwl**á**› (*ṮM* 38_4)
l-lŭwwla ‘first$^{f\ def}$’ (cf. SA الأولى ‹alåwlá› *ʔal-ʔūlā* ‘idem’)

(56)  

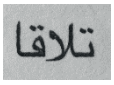
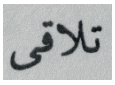
‹tns**a**› (*Ṣa.* 252^9) ‹tns**á**› (*ṮN* 48_1)
tənsa ‘you forget’ (cf. SA تنسى ‹tnsá› *tansā* ‘idem’)

(57)  

‹ytsn**a**› (*ṮN* 12_6) ‹ytsn**á**› (*WM* 39^{11})
yĭtsənna ‘he waits’
(cf. SA يتسنى ‹ytsná› *yatasannā* ‘it is feasible’, a morphological cognate)

(58)  

‹mš**a**› (*ḤB* 38^{12}) ‹mš**á**› (*Mu.* 14^3)
mša ‘he went’ (cf. SA مشى ‹mšá› *mašā* ‘idem’)

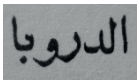
(59)  

‹tlaq**a**› (*ṮN* 20_{12}) ‹tlaq**á**› (*ḤD* 32^1)
tlaqa ‘he met’ (cf. SA تلاقى ‹tlaqá› *talāqā* ‘idem’)

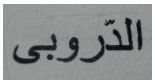
Since variation in (60) manifests itself in the graphy of the final *-a* marking the plural number, the graphies of this MA word need to be relativized to the graphy of its SA morphological, not lexical, cognates[111]. These cognates are SA plurals of the pattern $C_1uC_2\bar{u}C_3a^{tun}$, written with a final ة ‹ħ› or ه ‹h›. Hence, both ا ‹a› and ى ‹á› in this pair can be considered donor-defying graphies.

[111] Its SA lexical cognate is *darb^{un}* ‘lane, small street’, with the plural *durūb^{un}*.

(60)



‹aldrwb**a**› (*WM* 29[10])



‹ald·rwb**á**› (*Ṛḥ.* 198[7])

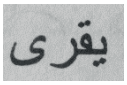
d-druba ‘lanes, small streets[def]’

(cf., for instance, السبوعة ‹alsbwʕh› *ʔas-subūʕa*[tu] ‘predatory animals[def]’)

Both graphies are sometimes used in verbs with SA cognates written with the word-final أ ‹å›, as in (61)-(63), and in non-verbs with SA cognates written with the word-final اء ‹ao›, as in (64) and (65). The graph ا ‹a›, being formally closer to the donor graphy, is much more frequent in such cases.

(61)

‹yqr**a**› (*MX* 59$_6$)

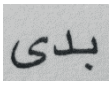


‹yqr**á**› (*Ṛḥ.* 129[2])

yəqṛa ‘he reads; he studies’ (cf. SA يقرأ ‹yqrå› *yaqraʔ*[u] ‘he reads’)

(62)

‹bd**a**› (*Ṣa.* 209$_4$)

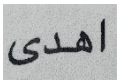


‹bd**á**› (*T̲N* 45[2])

bda ‘he began’ (cf. SA بدأ ‹bdå› *badaʔa* ‘idem’)

(63)

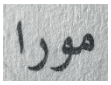
‹ahd**a**› (*MX* 18[1])



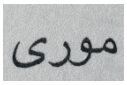
‹ahd**á**› (*T̲N* 41$_7$)

hda ‘calm down!’ (cf. SA اهدأ ‹ahdå› *ʔihdaʔ* ‘idem’)

(64)



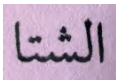
‹mwr**a**› (*Ḥ̣B* 14$_2$)



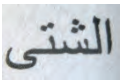
‹mwr**á**› (*T̲N* 54[11])

muṛa ‘behind, after’ (cf. SA من وراء ‹mn_wrao› *min warāʔⁱ* ‘from behind’)

(65)



‹alšt**a**› (*Ṣ̌X* 96[1])



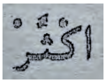
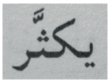
‹alšt**á**› (*Ṛḥ.* 158$_1$)

š-šta ‘rain[def]’ (cf. SA الشتاء ‹alštao› *ʔaš-šitāʔ*[u] ‘winter[def]’)

ا ‹a› ~ ي ‹y›

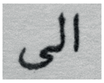
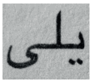
Word-initially / Restricted variation (ا ‹a› Rv ي ‹y›)

This variation manifests itself in graphic words representing words beginning with *i-* or *yV-*, especially in the prefix of the 3rd person non-feminine present tense, discussed under "ا ‹a› Av إ ‹ḁ›". The corpus reveals that either of these graphs can be used to mark either pronunciation, but in verbs, ي ‹y› is by far more frequently used, especially to mark the latter one. In order to rule out any influence of the phonetic environment on the pronunciation and graphy, every heterograph pair given below is composed of graphic words occurring in the same phonetic environment: in (66) the words are preceded by a word ending in a consonant, in (67) and (68) by words ending in a vowel.

(66)  

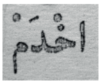
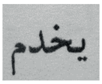
‹**a**k$^{\bullet}$ṯ$^{:a}$r$^{\bullet}$› (*Ḍʔ* 13^{5}) ‹**y**kṯ$^{:a}$r› (*MX* 107^{1})

ikəttər 'he multiplies' (cf. SA يكثّر ‹ykṯ$^{:}$r› *yukaṯṯir^{u}* 'idem')

(67)  

‹**a**lá› (*Ḍa.* $_{6}$44) ‹**y**lá› (*ṮN* 42$_{8}$)

ila 'if'

(68)  

‹**a**x$^{\bullet}$d^{a}m$^{\bullet}$› (*Ḍʔ* 13$_{4}$) ‹**y**xdm› (*XM* 91^{6})

yəxdəm 'he works' (cf. SA يخدم ‹yxdm› *yaxdim^{u}* 'he serves')

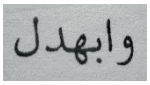
When the pronunciation of the prefix is *i-*, as in (66) and (67), the graphy ا ‹a› is adaptive phonetic, with the adaptation being endogenous and historically motivated. By contrast, the reason for ي ‹y› seems to be donor form-orientation if there is a SA cognate with ي ‹y›, as in (66). If there is no SA cognate that could suggest this spelling, as in (67), the graphy is analogical: The speller is aware that ي ‹y› is used in many MA words beginning with *i-*, e.g. verbs written in the donor form-oriented way, and uses it in a word beginning with *i-* in which, however, this spelling cannot be justified by donor form-orientation.

When the pronunciation is *yV-*, as in (68), the graphy ي ‹y› is adoptive phonetic, based on general donor graphy, while ا ‹a› can be interpreted as either donor-defying or analogical: The speller is aware that ا ‹a› is used in many cases of this MA verbal form and uses it in a verb which, however, is pronounced in a way not justifying this.

The analogical graphies may be additionaly induced by the relatively small auditive difference between the two pronunciations.

Word-medially / Restricted variation (ا ‹a› Rv ي ‹y›)
This variation manifests itself, rarely, in graphic words in which these graphs mark *i* and their word-medial position results from graphic prefixation. In the following examples, the prefixed element marks the conjunction *w-* ‘and’.

(69)



‹w**a**bhdl› (*Ṣa.* 215_4)

‹w**y**bhdl› (*Ḥm* 85_3)

w-ibəhdəl ‘and he humiliates’ (cf. SA ويبهدل ‹wybhdl› *wa-yubahdil*^{u} ‘idem’)

(70)

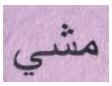
‹w**a**lá› (*Mn.* 42_{11})

‹w**y**lá› (*Ṯn* 60_5)

w-ila ‘and if’

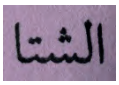
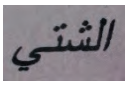
If the graph ا ‹a› marking the word-medial *i* is considered to be the remnant of a word-initial ا ‹a› (cf. ابهدل ‹abhdl› *ibəhdəl* ‘he humiliates’ and الى ‹alá› *ila* ‘if’ and see above for variation of this pair word-initially), this graphy is adaptive phonetic, with the adaptation being endogenous and historically motivated. The reason underlying the use of ي ‹y› in these examples is less obvious. First, one can assume that ي ‹y› is the remnant of its use word-initially. Then, in (69), the graphy ي ‹y› would be donor form-oriented, while in (70), which has no SA cognate, it would be explained as analogical graphy, cf. (66) and (67) for this variation word-initially. However, in both examples, the graphic word can also be viewed as an indivisible whole beginning with *wi-* (such as, for instance, ويل ‹wyl› *wil* ‘woe’), rather than as a graphically prefixed word beginning with *w-i-*. Then, the graphy ي ‹y› is adoptive phonetic, based on general donor graphy (cf. also “إ ‹ə› Rv ي ‹y›” word-medially).

Word-finally / Restricted variation (ا ‹a› Rv ي ‹y›)
This variation manifests itself in some words ending in *a* which also have spelling with ى ‹á› (discussed under “ا ‹a› Rv ى ‹á›”). It is, however, marginal because forms with a final ي ‹y› marking *a* are rare. Interestingly, all occurrences recorded in the corpus have SA cognates. Examples:

(71)  

‹mš**a**› (*ḤB* 38[12]) ‹mš**y**› (*Ṛḥ.* 168$_9$)

mša ‘he went’ (cf. SA مشى *mašā* ‘idem’)

(72)  

‹alšt**a**› (*ṢX* 96[1]) ‹alšt**y**› (*Ṛḥ.* 195$_8$)

š-šta ‘rain^{def}’ (cf. SA الشتاء ‹alštao› *ʔaš-šitāʔᵘ* ‘winter^{def}’)

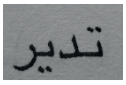
The graph ا ‹a› in (71) is donor-defying, while in (72) it is adoptive phonetic, based on general donor graphy. As for the use of ي ‹y›, it mirrors somehow the usage in SA, where it is, in some rare cases, used instead of ى ‹á› to mark final *a*. As such, it seems to be an instance of analogical graphy (see “ى ‹á› Fv ي ‹y›” for a detailed discussion).

Pseudo-variation

ا ‹a› Pv ت ‹t›

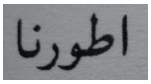
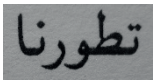
Word-initially

This intra-variety pseudo-variation occurs in graphic words representing verbs with the abstract prefix *t*-, usually inflectional, less often derivational, which assimilates to a stem-initial denti-alveolar stop – *d* in (73), *ṭ* in (74), and *ḍ* in (75) – and constitutes word-initial geminates with them. The graph ا ‹a› signals this word-initial gemination as a pseudo-prothesis and as such is donor principle-oriented, while ت ‹t›, marking the abstract prefix *t*-, is an instance of morphological graphy coincident with the donor graphy.

(73)  

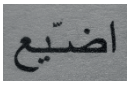
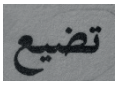
‹**a**dˑyr› (*ṮM* 29$_7$) ‹**t**dyr› (*ṮN* 23[7])

ddir (< *t-dir*) ‘she makes’ (cf. SA تدير ‹tdyr› *tudīrᵘ* ‘she manages, she directs’)

(74)  

‹**a**ṭwrna› (*XM* 109[1]) ‹**tt**ṭwrna› (*XM* 109[1])

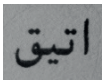
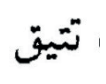
ṭṭŭwwəṛna (< *t-ṭŭwwəṛna*) ‘we developed’ (cf. SA تطورنا ‹ttṭwrna› *taṭawwarnā* ‘idem’)

(75)  

‹**a**ḍ̣yʕ› (*Ṛḥ.* 137[9]) ‹**t**ḍ̣yʕ› (*Mu.* 22$_9$)

ḍḍiʕ (< *t-ḍiʕ*) ‘it[f] gets lost’ (cf. SA تضيع ‹tḍ̣yʕ› *taḍ̣īʕ*[u] ‘idem’)

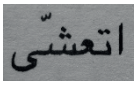
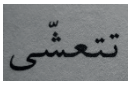
If the stem begins with *t*, the graphy ت ‹t› can also be interpreted as marking the initial geminate. As such, it is adoptive phonetic, based on general donor graphy. Example:

(76)  

‹**a**tyq› (*XM* 76$_4$) ‹**tt**yq› (*Ḥ̣M* 89$_1$)

ttiq (< *t-tiq*) ‘you trust’ (cf. SA تثق ‹tṯq› *taṯiq*[u] ‘idem’)

A different situation results from attaching this prefix to a consonant cluster begining with one of these stops. Then, a sequence of three consonants originates the pronunciation of which is “a complex phonological problem that needs to be examined in detail for each dialect” (Heath 2002: 205). Such clusters tend to be reduced in normal speech or split by inserting a *shwa* in a deliberately careful pronunciation. Consider the following example:

(77)  

‹**a**tʕšːá› (*Ṯ̲N* 43[7]) ‹**tt**ʕšːá› (*BB* 97$_9$)

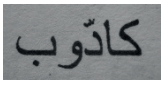
təʕšša (< *t-ʕšša*) ‘you have supper’ (cf. SA تتعشّى ‹ttʕšːá› *tataʕaššā* ‘idem’)

Here, ا ‹a› signals the pseudo-cluster *təʕ-* or, if the pronunciation [tʕəʃːa], with cluster simplification, is assumed, it functions as a pseudo-prothesis signalling a two-consonant cluster. In either case, the principles underlying particular graphies are the same as those indicated for (73)-(76).

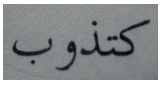
Word-medially

This intra-variety pseudo-variation manifests itself in graphic words which result from graphic prefixation (cf. pseudo-variation of this pair word-initially discussed above). For instance:

(78)



‹k**a**(d)ˑwb› (*Ṛḥ.* 13_5)
ka-ddub (< *ka-t-dub*)
‘you melt’



‹k**t**(ḏ)wb› (*Fʕ* 54_5)
‘she/it[f] melts’[112]

(cf. SA تذوب ‹tḏwb› *taḏūb*^{u} ‘idem’)

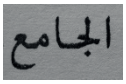
The graphy ت ‹t› in *B* marks the abstract inflectional affix *t-* and is thus morphological, coincident with the donor graphy. In this particular case, *t-* is normally assimilated to the stem-initial *d-*, the result being the geminate *dd*. The graph ا ‹a› in *A* is functionally indeterminate: It either marks *a* in the preverb *ka-* (adoptive phonetic graphy, based on general donor graphy) or is a remnant of the donor principle-oriented graphy ا ‹a› used pseudo-prothetically, cf. the unprefixed form ادّوب ‹adˑwb› *ddub* ‘she/it[f] melts’ (*Ṛḥ.* 191^3). Whatever be the case, the function of ا ‹a› is different than that of ت ‹t›[113], hence the pair is a case of a pseudo-variation.

ا ‹a› Pv ل ‹l›

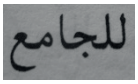
Word-initially

This infrequent intra-variety pseudo-variation manifests itself when the preposition *l(ə)-* ‘to, for’ is prefixed to a word with the definite article, which may be *l-* or a consonant resulting from the assimilation of *l-* to the stem-initial consonant, as in (79).

(79)



‹**a**lğamʕ› (*Ḍa.* 97^7)
lə-ž-žaməʕ ‘to the mosque’ (cf. SA للجامع ‹llğamʕ› *li-l-ğāmiʕ*^{i} ‘idem’)



‹**l**lğamʕ› (*MX* 15_{14})

---

[112] The fact that the affix *t-* signals the 2nd person in *A* and the 3rd person feminine in *B* does not affect pronunciation or morphology and can be disregarded here.

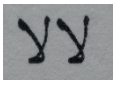
[113] Note that this pair only appears to be a pair of qualitative heterographs. In reality, it represents (contiguous) enantiodiacritic quantitative heterography: in *A* the augment is ا ‹a›, in *B* it is ت ‹t›.

The graphy ا ‹a›, quite infrequent, signals the pseudo-cluster *lə-ž-ž* and is thus donor principle-oriented (ل ‹l› marks the preposition *l(ə)-*, while ج ‹ǧ› marks the geminate resulting from the assimilation of the definite article to the stem-initial *ž*). In Form *B*, the graph ل ‹l› being in pseudo-variation with ا ‹a›, i.e. the first one in the graphic word, marks the preposition *l(ə)-*, its graphy being adoptive phonetic, based on general donor graphy, and coincident with the donor graphy (the second ل ‹l› marks the definite article, morphologically).

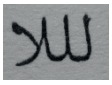
Word-medially

Intra-variety pseudo-variation of these graphs has been identified in one pair of graphic words, shown in (80). Both graphies, ا ‹a› and ل ‹l› are adoptive phonetic, based on general donor graphy but each of them has a different function: ا ‹a› marks *a*, while ل ‹l› marks *l*[114].

(80)



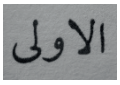
‹l**a**la› (*Fʕ* 49[8])
lalla ‘madam’



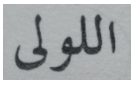
‹l**l**la› (*FL* 81[7])

Inter-variety pseudo-variation involving this pair is manifested in (81).

(81)



‹al**a**wlá› (*Ḥ D* 18[4])
l-lŭwwla ‘first[f def]’
(cf. SA الاولى ‹alawlá› *ʔal-ʔūlā* ‘idem’)



‹al**l**wlá› (*Ḥ B* 12$_{7}$)

In this pair, the graphy ا ‹a› is donor form-oriented, while the graphy ل ‹l› for *l* is adoptive phonetic, based on general donor graphy.

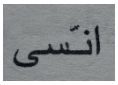
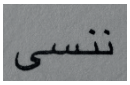
ا ‹a› Pv ن ‹n›

Word-initially

This intra-variety pseudo-variation manifests itself, marginally (in one source only), in the 1st person singular present tense verbs with a stem beginning with *n* or *r*. In the former case, the attaching of the abstract personal prefix *n-* to the stem results in the word-initial geminate *nn-*, while in the latter case, the prefix *n-* is

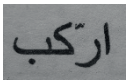
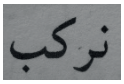
[114] This pair only appears to be a pair of qualitative heterographs. In reality, it represents (contiguous) enantiodiacritic quantitative heterography: in *A* the augment is ا ‹a›, in *B* it is ل ‹l›.

assimilable to *r* in less careful speech. The graphy ا ‹a›, used less frequently than ن ‹n› in this situation, but more often than could be expected, signals the word-initial geminate and is thus donor principle-oriented, while the graphy ن ‹n› is morphological, coincident with the donor graphy. Examples:

(82)  

‹**a**nːsá› (*Ṛḥ.* 10[12]) ‹**n**nsá› (*TN* 49[1])

nnsa [nːsa] ‘I forget’ (cf. SA ننسى ‹nnsá› *nansā* ‘we forget’)

(83)  

‹**a**rːkb› (*Ṛḥ.* 20[2]) ‹**n**rkb› (*MX* 33[14])

nrkəb [rːkəb] ‘I ride’ (cf. SA نركب ‹nrkb› *narkab*^{u} ‘we ride’)

GRAPH أ ‹å›

أ ‹å› ~ ا ‹a›

Word-initially / Asymmetrical variation (inverse) (أ ‹å› Av^{-1} ا ‹a›)

See ا ‹a› *Av* أ ‹å›

Word-medially / Restricted variation (أ ‹å› Rv ا ‹a›)

See ا ‹a› Rv أ ‹å›

Word-finally / Asymmetrical variation (inverse) (أ ‹å› Av^{-1} ا ‹a›)

See ا ‹a› *Av* أ ‹å›

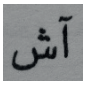
أ ‹å› ~ آ ‹ã›

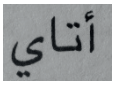
Word-initially / Asymmetrical variation (أ ‹å› Av آ ‹ã›)

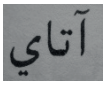
This variation manifests itself in graphic words in which the graphs mark the word-initial *a*. It is asymmetrical because أ ‹å› can always be substituted for آ ‹ã› in this function but not always can آ ‹ã› be substituted for أ ‹å›: Marking word-initial *u*, signalling initial clusters (pseudo-prothetically) or pseudo-clusters is a function of أ ‹å› (see examples under “ا ‹a› Av أ ‹å›”) but not آ ‹ã›. There are, for instance, no occurrences of *آستاد ‹ãstad› for *ʔŭstad* ‘professor’, *آحنا ‹ãḥna› for *ḥna* ‘we’ or *آبلي ‹ãbly› for *bəlli* ‘that’ (conjunction).

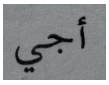
Both graphies marking *a* are adaptive phonetic, with the adaptation being endogenous and historically motivated. In (88), the MA word has a SA cognate, hence, the graphy ا ‹a› is coincident with the donor graphy.

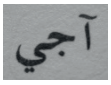
(84) 
‹**å**š› (*Ṣa.* 243[4])
aš ‘what’

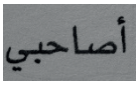

‹**ã**š› (*Bə.* 41[1])

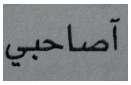
(85) 
‹**å**tay› (*ṮN* 26[6])
atay ‘tea’


‹**ã**tay› (*MX* 81_{12})

(86) 
‹**å**ğy› (*ṮM* 67_{6})
ažı ‘come!; hey!’


‹**ã**ğy› (*ḤB* 9_{3})

(87) 
‹**å**ṣaḥby› (*Bə.* 41[4])
a-ṣaḥb-i ‘my friend!’ (*a-* is a vocative particle)


‹**ã**ṣaḥby› (*Bə.* 74_{1})

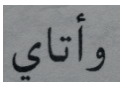
(88) 
‹**å**na› (*MX* 9_{2})
ana ‘I’ (cf. SA أنا ‹åna› *ʔana* ‘idem’)


‹**ã**na› (*Bə.* 45[3])

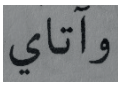
Word-medially / Restricted variation (أ ‹å› Rv آ ‹ã›)
This variation manifests itself only in graphically prefixed words[115]. It is analogous to variation between these graphs used word-initially, of which both are word-medial remnants. Both graphies marking *a* are adaptive phonetic, with the adaptation being endogenous and historically motivated. Example:

[115] Otherwise, these graphs may be meaning-differentiating, albeit marginally. One can think here of the pair: نأدي ‹nãdy› *nʔadi* ‘I hurt, I harm’ vs. نأدي ‹nådy› *nʔəddi* ‘I perform’. Although these graphic words have not been recorded in the corpus, they are expected on the basis of such graphic words as: نأديوا ‹nãdywa› *nʔadiw* ‘we harm’ (*MX* 169_{7}) and أنأدّي ‹ån·åd·y› *nʔəddi* ‘I perform’ (*MX* 123_{6}).

(89)



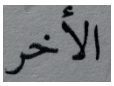
‹w**å**tay› (*MX* 49$_4$)
w-atay ‘and tea’



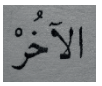
‹w**ã**tay› (*MX* 81)

In the case of graphic words with SA cognates, the adaptive phonetic graphy can be relativized to the donor graphy. Thus, the graphy آ ‹ã› in (90) and أ ‹å› in (91) are coincident with the donor graphy, while أ ‹å› in (90) and آ ‹ã› in (91) are donor form-defying.

(90)



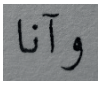
‹al**å**xr› (*LŠ* 96$_3$)
l-axŭṛ ‘other^{def}’ (cf. SA الآخر ‹alãxr› *ʔal-ʔāxar^{u}* ‘idem’)



‹al**ã**x^{u}r̓› (*Mm.* 111$_2$)

(91)

‹w**å**na› (*WM* 31^6)
w-ana ‘and I’ (cf. SA وأنا ‹wåna› *wa-ʔana* ‘idem’)



‹w**ã**na› (*DB* 142$_3$)

أ ‹å› ~ إ ‹ą̊›

Word-initially / Restricted variation (أ ‹å› Rv إ ‹ą̊›)
This variation manifests itself in words with word-initial *i*, *ĭ* or *y*. The graphy إ ‹ą̊› is adaptive phonetic, with the adaptation being endogenous and historically motivated. If there is a SA cognate, the graphy أ ‹å›, less frequently used, is donor form-oriented. Examples:

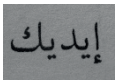
(92)

‹**å**yam› (*ḤD* 25^5)
ĭyyam ‘days’ (cf. SA أيام ‹åyam› *ʔayyām^{un}* ‘idem’)

‹**ą̊**y˙am› (*T̲N* 58$_2$)

(93)

‹**å**ydy[h]› (*Ṣa.* 220$_6$)
yĭddi[-*h*] ‘his hands’



‹**ą̊**ydy[k]› (*Ṣa.* 227$_2$)
yĭddi[-*k*] ‘your hands’
(cf. SA أيديه ‹åydyh› *ʔaydī-hi* ‘his hands^{pl}’)

If there is no SA cognate that could be considered to be serving as a model, the graphy أ ‹å› is analogical, as in (94).

(94)

‹**å**ymta› (*ḤD* 19_8)
imta ‘when?’

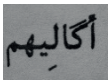
‹**ą**ymta› (*Ṣa.* 238^4)

The speller of Form *A* is aware that the word-initial *ĭ* can be written as أ ‹å› in some words, for instance *ĭyyam* (donor-oriented graphy), and (s)he applies this graphy to *imta*, a word which begins with *i*, a sound phonetically similar, but has no SA cognate that would justify this graphy.

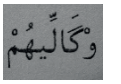
أ ‹å› ~ و ‹w›

Word-initially / Restricted variation (أ ‹å› Rv و ‹w›)
This variation manifests itself in the graphy of *u-* ‘and’, the phonetic variant of the conjunction used before consonants (with *w-* being used before vowels, Harrell 1962: 212; cf. also Aguadé & Benyahia 2005: 9 and Durand 2004: 72, who, in addition, notes instability affecting the use of these two sounds[116]). The graphy أ ‹å› used in this function (only in *Ḍa.*) is adaptive phonetic, with the adaptation being endogenous and historically motivated, while و ‹w› is donor form-oriented. Examples:

(95)

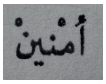


‹**å**ǵal^{i}yhm› (*Ḍa.* 47_4)
u-gal-li-hŭm ‘and he said to them’
(cf. SA وقال لهم ‹wqal_lhm› *wa-qāla la-hum* ‘idem’)

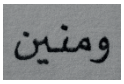


‹**w**ˑǵ^{a}al$^{:i}$yh^{u}mˑ› (*Mm.* 116_1)

(96)



‹**å**^{u}mˑnynˑ› (*Ḍa.* 72_4)
u-mnayn ‘and from where’
(cf. SA ومن أين ‹wmn_åyn› *wa-min ʔayn* ‘idem’)



‹**w**mnyn› (*Bə.* 44_9)

---

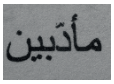
[116] Indeed, there are many pairs like *udən ~ wdən* ‘ear’, *ulad ~ wlad* ‘sons, children’, *uṣəḷ ~ wṣəḷ* ‘he arrived’; see Aguadé & Benyahia (2005: 153-154) for more.

In contrast to the similar case of ا ‹a›/إ ‹ạ› and ي ‹y› marking the vowel *i*- and the semivowel *y*-, respectively, where the correspondence between graphs and sounds was sometimes violated (see example 32 under "ا ‹a› Av إ ‹ạ›" word-initially), in the present case there is consistency: No instances of the semivowel *w*- being marked as أ ‹å›, *typically* marking vowels (e.g. *أُأنا ‹åᵘana› for *w-ana* 'and I'), have been recorded in the corpus.

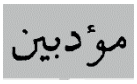
أ ‹å› ~ ؤ ‹ẘ›

Word-medially / Restricted variation (أ ‹å› Rv ؤ ‹ẘ›)
This variation manifests itself, marginally, in graphic words representing borrowings from SA in which they mark *ʔ*. For example:

(97)



‹mådˑbyn› (*Ṛḥ.* 97[6])



‹mẘdbyn› (*ḤM* 87$_9$)

mʔəddbin 'well-mannered[pl]'
(cf. SA مؤدّبين ‹mẘdˑbyn› *muʔaddabīn*[a] 'well-mannered[pl obl]')

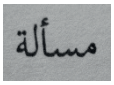
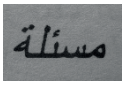
The graphy ؤ ‹ẘ› is donor form-oriented, whereas أ ‹å› could seem to be donor form-defying, but can also be interpreted as donor principle-oriented, i.e. being a result of applying to MA pronunciation the SA orthographic rule according to which a glottal stop preceded by a consonant and followed by *a* is marked as أ ‹å› (MA *ə* would be treated like *a* in this case) (cf. also "أ ‹å› Rv ئ ‹ỷ›" word-medially).

أ ‹å› ~ ى ‹á›

{All possible positions (word-finally) / Restricted variation (أ ‹å› Rv ى ‹á›)}
This variation is expected to manifest itself in graphic words representing verbs with SA cognates in which *ʔ* is the final radical. The donor form-oriented graphy أ ‹å› has been observed in such verbs as يبرأ ‹ybrå› *yəḇṛa* 'he gets cured' (*BB* 70[7]) and يتوضأ ‹ytwḍå› *yĭtwŭḍḍa* 'he performs his ablutions' (*Ḍa.* 97[7]), while their heterographs with ى ‹á›, adoptive phonetic graphy based on general donor graphy, are missing. By contrast, for the verbs يقرى ‹yqrá› (*Ṛḥ.* 129[2]) *yəqṛa* 'he reads; he studies' and نبدى ‹nbdá› (*T̲N* 49[5]) *nəbda* 'I begin', the graphy ى ‹á›, i.e. adoptive phonetic graphy based on general donor graphy, is recorded, while no cases of donor form-oriented graphy أ ‹å› have been encountered.

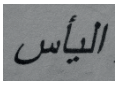
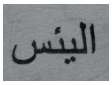
أ ‹å› ~ ئ ‹ŷ›

Word-medially / Restricted variation (أ ‹å› Rv ئ ‹ŷ›)
This variation manifests itself, marginally, in graphic words in which these graphs mark *ʔ*. The graphy أ ‹å› is donor form-oriented, whereas ئ ‹ŷ›, a marginal graphy, can be donor form-defying, as in (98).

(98)  

‹msålħ› (*Mn.* 73_3) ‹msŷlħ› (*Ṛḥ.* 149^8)
məsʔala ‘issue’ (cf. SA مسألة ‹msålħ› *masʔala*^{tun} ‘idem’)

It can also be donor principle-oriented, as in (99), in which probably the SA orthographic rule requiring a glottal stop preceded by *i* and followed by a consonant to be marked as ئ ‹ŷ› was applied to MA pronunciation (cf. also “أ ‹å› Rv ؤ ‹ẘ›” word-medially).

(99)  

‹alyås› (*Ṛḥ.* 83_7) ‹alyŷs› (*Ṛḥ.* 201^{11})
l-yīʔs ‘despair^{def}’ (cf. SA اليأس ‹alyås› *ʔal-yaʔs*^{u} ‘idem’)

{Word-finally / Restricted variation (أ ‹å› Rv ئ ‹ŷ›)}
This variation expected to manifest itself in graphs marking word-final *ʔ* in some borrowings from SA used rarely in MA texts, would be marginal. The graph أ ‹å› in خاطأ ‹xaṭå› *xaṭəʔ* ‘wrong’(*MX* 187^{12}) can be interpreted as donor principle-oriented graphy (assuming that the speller applied the SA orthographic rule on the spelling of word-final *-aʔ* to MA *-əʔ*). By contrast, the graphy ئ ‹ŷ› in هادئ ‹hadŷ› *hadəʔ* ‘calm’ (*Ṛḥ.* 101_{10}) is donor form-oriented.

أ ‹å› ~ ء ‹o›

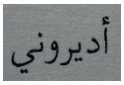
{Word-finally / Restricted variation (أ ‹å› Rv ء ‹o›)}
Like the preceding one, this expected variation would be marginal. Two graphic words with donor form-defying graphy ء ‹o› have been recorded, الخطء ‹alxṭo› *l-xaṭaʔ* ‘mistake^{def}’ (*Ṛḥ.* 157^4) and تجرّء ‹tğr·o› *džərrəʔ* ‘he dared’ (*Ṛḥ.* 28_{10}), while the respective donor form-oriented graphies, الخطأ ‹alxṭå› and تجرأ ‹tğrå› (cf. SA الخطأ ‹alxṭå› *ʔal-xaṭaʔ*^{u} ‘idem’ and تجرأ ‹tğrå› *tağarraʔa* ‘idem’) are expected. Both these words are borrowings from SA and are rather unusual in MA.

Pseudo-variation

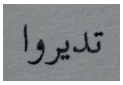
أ ‹å› Pv ت ‹t›

Word-initially
This intra-variety pseudo-variation occurs in graphic words in which أ ‹å› signals, as a pseudo-prothesis, word-initial gemination, cluster or pseudo-cluster resulting from attaching the abstract prefix *t-*. Example:

(100)



‹**å**dyrw[ny]› (*Ṣa.* 232[3])
ddiru[*-ni*] (< *t-diru*[*-ni*])
'you[pl] make me'



‹**t**dyrw(a)› (*MḤ* 68[5])
ddiru (< *t-diru*)
'you[pl] make'

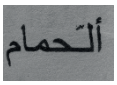
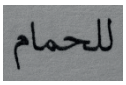
(cf. SA تديروا ‹tdyrwa› *tudīrū* 'that you[pl] manage, that you[pl] direct')

The graphy ت ‹t›, marking the prefix, is morphological, coincident with the donor graphy, while أ ‹å›, used rather infrequently (in *Ṣa.*), is donor principle-oriented, although SA uses ا ‹a›, not أ ‹å›, in this function. The use of أ ‹å› can be explained as analogical graphy: The speller is aware that أ ‹å› and ا ‹a› are interchangeable in MA in many word-initial contexts (e.g. when marking *a*) and (s)he uses أ ‹å› to signal pseudo-prothesis, even though this is usually made by means of ا ‹a›.

أ ‹å› Pv ل ‹l›

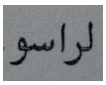
Word-initially
This intra-variety pseudo-varation occurs (mostly in *Ṛḥ.*) when the pseudo-prothetic graph أ ‹å› is used to signal a (pseudo-)cluster or a geminate resulting from prefixing the preposition *l(ə)-* 'to, for' to a word beginning with a consonant or consonant cluster and, as in the case of (102), its assimilation[117]. The graphy أ ‹å› is interpretable as donor principle-oriented or analogical (cf. discussion of the preceding pair). The graphy ل ‹l› is adoptive phonetic based on general donor graphy in (101), but morphological in (102), where the preposition undergoes assimilation. In both cases, it is coincident with the donor graphy.

[117] Form *B* in (102) is assumed to have two pronunciations: *ṛ-ṛaṣ-u*, with assimilation, and *l-ṛaṣ-u*, without it, the latter not concerning us here.

(101)  

‹**å**l·ḥmam› (*Ṛḥ.* 37[6]) ‹**l**lḥmam› (*Mn.* 29$_{6}$)

lə-l-ḥəmmam ‘to the bathroom’ (cf. SA للحمام ‹llḥmam› *li-l-ḥammām*^{i} ‘idem’)

(102)  

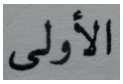
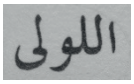
‹**å**rasw› (*Ṣa.* 237[3]) ‹**l**rasw› (*ʕD* 196$_{12}$)

ṛ-ṛaṣ-u (< *l-ṛaṣ-u*) ‘to himself’, lit. ‘to his head’

(cf. SA لرأسه ‹lråsh› *li-raʔsi-hi* ‘to his head’)

Word-medially

In this marginal inter-variety pseudo-variation (like in the case of “ا ‹a› Pv ل ‹l›”), the graphy أ ‹å› is donor form-oriented (أ ‹å› marks *ʔ* in the SA cognate), while the graph ل ‹l›, marking *l*, is an adoptive phonetic graphy, based on general donor graphy. Example:

(103)  

‹al**å**wlá› (*Ḍa.* 66[10]) ‹al**l**wlá› (*ḤB* 12$_{7}$)

l-lŭwwla ‘first$^{f\ def}$’ (cf. الأولى ‹alåwlá› *ʔal-ʔūlā* ‘idem’)[118]

أ ‹å› Pv ن ‹n›

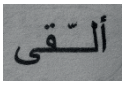
Word-initially

This intra-variety pseudo-variation, occurring exceptionally, in *Ṛḥ.*, is similar to the one described under “أ ‹å› Pv ل ‹l›” word-initially. In less careful speech, the prefix of the first person singular present tense, *n*-, can undergo assimilation to a stem-initial *l*, which results in a word-initial geminate *ll*[119]. In the pair below,

[118] A dubious instance of this pseudo-variation has been encountered in the pair of graphic words with the meaning ‘earth^{def}’. Whereas one of them, اللرض ‹allrḍ› (e.g. *ḤB* 8$_{4}$), has univocal graphy and is pronounced *l-lərḍ*, it is not sure if the other one, الأرض ‹alårḍ› (e.g. *Ṛḥ.* 144$_{9}$), represents the same pronunciation or a different one, viz. *l-ʔəṛḍ*. According to Cantineau (1950: 2016) and Kjamilev (1968: 24), *l-ʔəṛḍ* is used in the north of Morocco, while *l-lərḍ* is characteristic of the south. Heath (2002: 180), by contrast, claims that the glottal stop in *l-ʔəṛḍ* tends to be restored “under literary influence”. In view of this uncertainty, the pair is not taken into consideration here.

[119] Form *B* is assumed to have two pronunciations: *llqa*, with assimilation, and *nlqa*, without it, the latter not concerning us here.

this geminate is signalled by the pseudo-prothetic أ ‹å›, a donor principle-oriented graphy. The graphy ن ‹n› represents the abstract prefix and is thus morphological, coincident with the donor graphy. Example:

(104)  

‹**å**lˑqá› (*Ṛḥ.* 145[9]) ‹**n**ᵃlˑqᵃá› (*MX* 65$_{10}$)

llqa (< *n-lqa*) ‘I find, I meet’ (cf. SA نلقى ‹nlqá› *nalqā* ‘we find, we meet’)

GRAPH آ ‹ã›

آ ‹ã› ~ ا ‹a›

Word-initially / Asymmetrical variation (inverse) (آ ‹ã› Av^{-1} ا ‹a›)

See ا ‹a› Av آ ‹ã›

Word-medially / Restricted variation (آ ‹ã› Rv ا ‹a›)

See ا ‹a› Rv آ ‹ã›

آ ‹ã› ~ أ ‹å›

Word-initially / Asymmetrical variation (inverse) (آ ‹ã› Av^{-1} أ ‹å›)

See أ ‹å› Av آ ‹ã›

Word-medially / Restricted variation (آ ‹ã› Rv أ ‹å›)

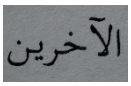
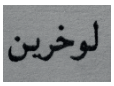
See أ ‹å› Rv آ ‹ã›

Pseudo-variation

آ ‹ã› Pv و ‹w›

Word-medially

In this marginal inter-variety pseudo-variation, exemplified in (105), آ ‹ã› copies the SA graph marking the sequence *ʔā* and is thus donor form-oriented, while the graphy و ‹w›, marking *ŭ*, is adoptive phonetic, based on general donor graphy.

(105)  

‹(a)l**ã**xryn› (*T̲M* 65[6]) ‹l**w**xryn› (*XM* 85_5)
l-ŭx̣rin ‘others[def]’ (cf. SA الآخرين ‹alãxryn› *ʔal-ʔāxarīn*[a] ‘others[def obl])

GRAPH ¦ ‹ą̊›

¦ ‹ą̊› ~ ا ‹a›

Word-initially / Asymmetrical variation (inverse) (¦ ‹ą̊› Av[-1] ا ‹a›)
See ا ‹a› Av ¦ ‹ą̊›

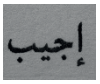
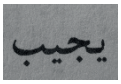
Word-medially / Asymmetrical variation (inverse) (¦ ‹ą̊› Av[-1] ا ‹a›)
See ا ‹a› Av ¦ ‹ą̊›

¦ ‹ą̊› ~ أ ‹å›

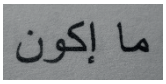
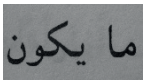
Word-initially / Restricted variation (¦ ‹ą̊› Rv أ ‹å›)
See أ ‹å› Rv ¦ ‹ą̊›

¦ ‹ą̊› ~ ي ‹y›

Word-initially / Restricted variation (¦ ‹ą̊› Rv ي ‹y›)
This variation manifests itself in graphic words in which they mark word-initial *i-* and *y-*. For a discussion, see the variation “ا ‹a› Rv ي ‹y›”, to which it is analogous. In what follows, the variation is illustrated with pairs of verbs preceded by the same phonetic environment in order to rule out its influence on the pronunciation and graphy. The most conspicuous category of words affected by this variation are the 3rd person non-feminine present tense verbs, for instance:

(106)  

‹**ą̊**ğyb› (*Ḍa.* 19_7) ‹**y**ğyb› (*MX* 91_3)
ižib ‘he brings’ (after a consonant)
(cf. SA يجيء ‹yğyo› *yağīʔᵘ* ‘he comes’, morphological cognate)

(107)  

‹ma_**ą**kwn› (*Ṛḥ.* 156[6]) ‹ma_**y**kwn› (*MX* 17$_9$)
ma ikun ‘he/it[m] is not’ (after a vowel)
(cf. SA لا يكون ‹la_ykwn› *lā yakūn*^{u} ‘idem’)

(108)  

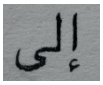
‹**ą**qra› (*Ḍ?* 212[8]) ‹**y**qra› (*BT* 117$_7$)
yəqṛa ‘he reads’ (after a vowel)
(cf. SA يقرأ ‹yqrå› *yaqraʔ*^{u} ‘idem’)

In the case of verbs pronounced with initial *i*-, ¦ ‹ą› is used phonetically as a result of historically motivated endogenous adaptation. The reason for the use of ي ‹y› seems to be donor form-orientation: in SA, this prefix is always written in this way. When the pronunciation is *yV*-, the graphy ي ‹y› is adoptive phonetic, based on general donor graphy, while ¦ ‹ą› can be interpreted as donor form-defying or analogical: The speller is aware that ¦ ‹ą› is used in some cases to mark this MA verbal prefix and uses it here even though the prefix is pronounced in a way not justifying this. In the case of both pronunciations, but especially *yV*-, the graphy ي ‹y› is by far more frequent (¦ ‹ą› being used only in three sources in the corpus: *Ṛḥ.*, *Ḍa.* and *Ḍ?*).

This variation also manifests itself in words for which only one pronunciation, *i*-, is given in dictionaries. Examples:

(109)  

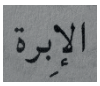
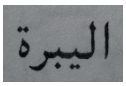
‹**ą**tym› (*Ṛḥ.* 178[2]) ‹**y**tym› (*MḤ* 42[12])
itim ‘orphan’ (cf. SA يتيم ‹ytym› *yatīm*^{un} ‘idem’)

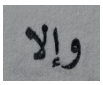
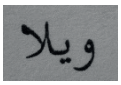
(110)  

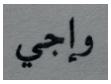
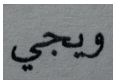
‹**ą**lá› (*TN* 8[13]) ‹**y**lá› (*ḌK* 8$_2$)
ila ‘if’

In both pairs, ¦ ‹ą› is used phonetically as a result of historically motivated endogenous adaptation. The graphy ي ‹y› is donor form-oriented in (109), while in (110) it must be explained as analogical graphy: The speller is aware that word-initial *i* is spelt ي ‹y› in some MA words, e.g. *itim* (principle of donor form-orientation), and (s)he applies this to *ila*, although it has no SA cognate that would justify this.

Word-medially / Restricted variation (ـِ ‹ą› Rv ـي ‹y›)
This variation mirrors the one discussed for this pair word-initially. It is also analogous to the variation ـا ‹a› Rv ـي ‹y› word-medially. It manifests itself in graphic words in which the word-medial position of ـِ ‹ą› and ـي ‹y› results from graphic prefixation. Examples:

(111)  
‹al**ą**^{i}brħ› (*MX* 114$_{15}$) ‹al**y**brħ› (*MX* 181$_{4}$)
l-ibra ‘needle^{def}’ (cf. SA الإبرة ‹alą̣brħ› *ʔal-ʔibra*^{tu} ‘idem’)

(112)  
‹w**ą**la› (*ḤB* 43$_{11}$) ‹w**y**la› (*ṮM* 9$_{10}$)
w-ila ‘and if’

(113)  
‹w**ą**ğy› (*Ḍa.* 99^{8}) ‹w**y**ğy› (*ṢX* 95$_{8}$)
w-iži ‘and he comes’ (cf. SA ويجيء ‹wyğyo› *wa-yağīʔ*^{u} ‘idem’)

This situation is analogous to the one discussed for this pair word-initially without graphic prefixes (cf. the preceding variation). The graph ـِ ‹ą› can be considered to be the remnant of a word initial ـِ ‹ą› (cf. إبرة ‹ąbrħ› *ibra* ‘needle’, إلا ‹ąla› *ila* ‘if’, إجي ‹ąğy› *iži* ‘he comes’). As such, it is used phonetically as a result of historically motivated endogenous adaptation. However, if the graphic word is not viewed as graphically prefixed but as indivisible (such as, for instance, ويل ‹wyl› *wil* ‘woe’), the graphy ـي ‹y› in (111)-(113) can be interpreted as adoptive phonetic, based on general donor graphy. In (113), for which a SA cognate exists, it is coincident with the donor graphy.

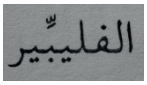
GRAPH ب ‹b›

ب ‹b› ~ پ ‹p›

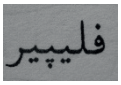
All possible positions / Asymmetrical variation (ب ‹b› Av پ ‹p›)
This variation manifests itself when these graphs mark *p*, a sound for which the SA alphabet has no special letter. In order to represent it, MA spellers sometimes use ب ‹b›, the graph normally marking *b*, its voiced counterpart, in both SA and

MA, and thus resort to adaptive phonetic graphy, with the adaptation being endogenous and phonetically motivated. In some cases, a *šadda*, normally marking the lengthening, or gemination, of a consonant, is written over ب ‹b› to mark that *p* is intended, as in Form *A* in (114). Marking this sound by means of the graph پ ‹p›, i.e. ب ‹b› with three dots, not included in the SA alphabet (but marking *p* in, for instance, Persian), also means resorting to adaptive phonetic graphy, but in this case the adaptation is exogenous. Below are given examples of word-medial and word-final occurrence. This variation is also expected to manifest itself word-initially.

(114)

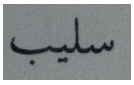


‹[al]fly**b**$^{\cdot i}$yr› (*TM* 43$_9$)
[*lə-*]*flipər* ‘flipper$^{\text{def}}$’

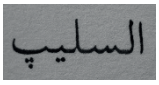


‹fly**p**yr› (*TM* 47^9)
flipər ‘flipper’

(115)



‹sly**b**› (*Fʕ* 18^2)
slip ‘briefs’



‹[al]sly**p**› (*TM* 17$_4$)
[*s-*]*slip* ‘briefs$^{\text{def}}$’

GRAPH پ ‹p›

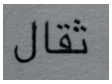
پ ‹p› ~ ب ‹b›

All possible positions / Asymmetrical variation (inverse) (پ ‹p› Av^{-1} ب ‹b›)
See ب ‹b› Av پ ‹p›

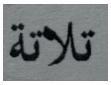
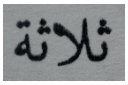
GRAPH ت ‹t›

ت ‹t› ~ ث ‹ṯ›

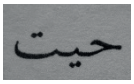
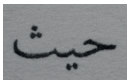
All possible positions / Asymmetrical variation (ت ‹t› Av ث ‹ṯ›)
This variation manifests itself, quite frequently, in words in which these graphs mark *t*. The graphy ت ‹t› is adoptive phonetic, based on general donor graphy. In words which have SA cognates with *t*, as in (119) and (120), it is, in addition, coincident with the donor graphy. The graphy ث ‹ṯ› (cf. Aguadé 2006: 257-258) is donor form-oriented in words which have SA cognates with *ṯ*, for instance:

(116)  

‹**t**qal› (*Mu.* 44_5) ‹**ṯ**qal› (*ṮN* 14_1)
tqal ‘it^{m} became heavy’ (cf. SA ثقل ‹ṯql› *ṯaqula* ‘idem’)

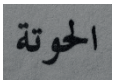
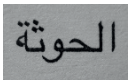
(117)  

‹**t**la**t**ḥ› (*DB* 211^{6}) ‹**ṯ**lā**ṯ**ḥ› (*LŠ* 97^{2})
tlata ‘three’ (cf. SA ثلاثة ‹ṯlāṯḥ› *ṯalāṯa*^{tun} ‘idem’)

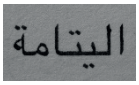
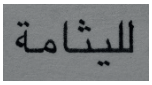
(118)  

‹ḥy**t**› (*ṮM* 11^{11}) ‹ḥy**ṯ**› (*Ṣa.* 245^{1})
ḥit ‘because’ (cf. حيث ‹ḥyṯ› *ḥayṯu* ‘idem’)

If the MA word has no SA cognate with a corresponding *ṯ*, the graphy ث ‹ṯ› is explainable in terms of analogical graphy: The speller is aware that *t* is spelt as ث ‹ṯ› in many MA words (by virtue of donor form-orientation) and (s)he uses it although the word to be written has no SA cognate with the corresponding *ṯ* that would justify this graphy. Examples:

(119)  

‹alḥw**t**ḥ› (*ḤB* 42^{11}) ‹alḥw**ṯ**ḥ›(*ṮN* 12^{3})
l-ḥuta ‘fish^{def}’ (cf. SA الحوت ‹alḥwt› *ʔal-ḥūt*^{u} ‘idem; whale^{def}’)

(120)  

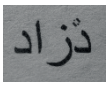
‹(a)ly**t**amḥ› (*ṮN* 42_{10}) ‹[l]ly**ṯ**amḥ› (*ṮN* 53^{11})
l-itama ‘orphans^{def}’ [*lə-*]*l-itama* ‘to orphans^{def}’
(cf. SA اليتامى ‹alytamá› *ʔal-yatāmā* ‘orphans^{def}’)

ت ‹t› ~ د ‹d›

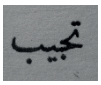
Word-initially / Restricted variation (ت ‹t› Rv د ‹d›)
This variation manifests itself in graphic words representing words in which the abstract affix *t-*, whether derivational or inflectional, is pronounced *d* due to assimilation. In such cases, the graphy ت ‹t› is morphological, whereas د ‹d›, rather infrequent, is adoptive phonetic, based on general donor graphy. Examples:

(121) 

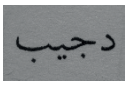
‹**t**zad› (*Ṛḥ.* 122_8)
dzad (< *t-zad*) ‘he was born’



‹**d**˙zad› (*Ṛḥ.* 102_5)

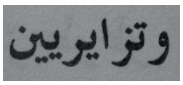
(122) 

‹**t**ğyb› (*ṢX* 92_8)
džib (< *t-žib*) ‘you bring’

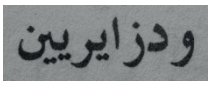


‹**d**ğyb› (*Ṣa.* 253^1)

Word-medially / Restricted variation (ت ‹t› Rv د ‹d›)
Although ت ‹t› and د ‹d› are bound by variation word-initially, manifesting itself in verbs (see preceding paragraph), no pair of graphically prefixed verbs illustrating this variation word-medially could be identified in the corpus. The reason is that word-medially, the affix *t-* is always written morphologically as ت ‹t›, e.g. *ka-džib* (< *ka-t-žib*) ‘you bring’ is spelt كتجيب ‹ktğyb›, while phonetic graphies (كدجيب ‹kdğyb›) have not been recorded. An isolated occurrence of this variation in the word-medial position involves a noun with a graphically prefixed conjunction *u-* ‘and’:

(123) 

‹w**t**zayryyn› (*Mu.* 43^{13})
u-dzayriyin ‘and Algerians’

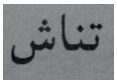
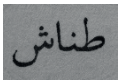


‹w**d**zayryyn› (*Mu.* 33_3)

The graphy د ‹d› is adoptive phonetic, based on general donor graphy, whereas ت ‹t› seems to be an instance of analogical graphy: The speller is aware that MA *d* is sometimes written as ت ‹t› (by virtue of the morphological principle) and uses it in this word despite the fact that *d* is part of its root, not an affix, and the graphy ت ‹t› cannot be justified by this principle.

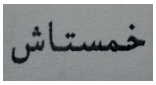
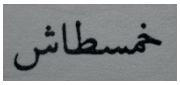
ت ‹t› ~ ط ‹ṭ›

All possible positions / Restricted variation (ت ‹t› Rv ط ‹ṭ›)
This variation manifests itself, rather marginally, in graphic words in which these graphs mark *ṭ*. It has been recorded word-initially and word-medially. Examples:

(124)  

‹**t**naš›(*MX* 79$_4$) ‹**ṭ**naš› (*FʕḤ* 57$_6$)

ṭnaš ‘twelve’ (cf. SA اثنا عشر ‹aṯna_ʕšr› *ʔiṯnā ʕašara* ‘idem’)

(125)  

‹xms**t**aš› (*Ḍa.* 98[1]) ‹xms**ṭ**aš› (*ḤD* 27[4])

xəmṣṭaš ‘fifteen’ (cf. SA خمسة عشر ‹xmsḥ_ʕšr› *xamsata ʕašara* ‘idem’)

The graphy ط ‹ṭ› is adoptive phonetic, based on general donor graphy. The use of ت ‹t› seems to be donor-defying in (124)[120], while in (125), it is donor form-oriented[121]. The variation is marginal due to the marginal use of ت ‹t› in this function.

For the word-final position, no heterograph pairs can be identified because heterographs with a word-final ت ‹t› have not been recorded in the corpus for such graphic words as كونط ‹kwnṭ› *kunṭ* (from French *compte*) ‘bank account’ (*Mu.* 19[11]) or الطواليط ‹alṭwalyṭ› *ṭ-ṭwaliṭ* (from French *toilette*) ‘toilet’ (*Mu.* 37[2]). However, since the variation between ت ‹t› and ط ‹ṭ› occurs when these graphs mark *ṭ* being part of the lexical root, it can be expected to occur in all positions, also word-finally. Hence, the missing heterographs كونت ‹kwnt› and التواليت ‹altwalyt› are expected forms.

ت ‹t› ~ ة ‹ħ›

All possible positions (word-finally) / Restricted variation (ت ‹t› Rv ة ‹ħ›)

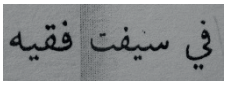
This variation manifests itself in nouns, mostly feminine, in the construct state. Both graphs mark the word-final *t* which replaces, or, in less numerous cases, is attached to, the final *-a* which the noun has in the free state (i.e. when it is not followed by a genitive attribute). The most frequent graphy of this *t* is ة ‹ħ› (the usual graphy in SA, cf. “ه ‹h› Rv ة ‹ħ›”), which may be characterized as adoptive phonetic, based on general donor graphy and coincident with the donor graphy. As for ت ‹t›, this graphy was used in CA, for instance, in the Quran, where the word *ʔimraʔat-* ‘woman, wife’ in the construct state is spelt in this manner if it is fol-

[120] Surprisingly, the expected form with donor form-oriented graphy ث ‹ṯ›, تناش ‹ṯnaš›, has not been recorded in the corpus.

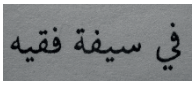
[121] Note that the correspondent graph in the SA cognate is not ت ‹t› but ة ‹ħ›. Since ة ‹ħ› cannot be used word-medially, its word-medial counterpart is used in the MA graphic word. This strictly positional issue can be disregarded here.

lowed by the husband's name (al-ʔAsmar 1988: 223), e.g. امرأت نوح ‹amråt_nwḥ› *ʔimraʔat^{u} nūḥ^{in}* 'Noah's wife' (surah 66, 10). However, it is rather unlikely that this use of ت ‹t› in MA should be based on the Quranic practice. It seems rather to be a donor form-defying graphy. Examples:

(126)



‹fy_syf**t**_fqyh› (*Ṣa.* 244[1-2])

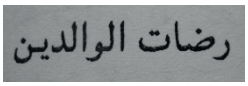


‹fy_syf**ħ**_fqyh› (*Ṣa.* 241$_{7}$)

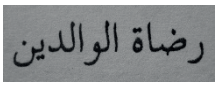
f-ṣift fqih 'in the form of a learned man' (cf. *ṣifa* in the free state)

(cf. SA في صفة فقيه ‹fy_ṣfħ_fqyh› *fī ṣifa^{ti}faqīh^{in}* 'in the capacity as legist')

(127)



‹rḍa**t**_alwaldyn› (*Ḍa.* 101$_{3}$)

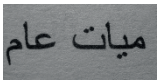


‹rḍa**ħ**_alwaldyn› (*MḤ* 26$_{9}$)

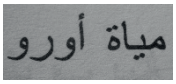
ṛḍat l-waldin 'the parents' joy' (cf. *ṛḍa* in the free state)

(cf. SA رضا الوالدين ‹rḍa_alwaldyn› *riḍā l-wālidayn^{i}* 'idem')

(128)



‹mya**t**[_ʕam]› (*Ṛḥ.* 161$_{2}$)



‹mya**ħ**[_åwrw]› (*Ṛḥ.* 126[3])

myat [*ʕam*]

'a hundred [years]'

myat [*oṛo*] (cf. *mya* in the free state)

'a hundred [euros]'

(cf. SA مئة عام ‹mỷħ_ʕam› *miʔat^{u} ʕām^{in}* 'a hundred years')

Since ة ‹ħ› is in variation with ه ‹h›, it might seem that also variation between ت ‹t› and ه ‹h› is possible. This is, however, not the case because, as the corpus findings show, forms with ه ‹h›, e.g. كلمه ‹klmh› 'word' (*MX* 170[4]), are always pronounced with a final *a* (*kəlma*, in the present example) and are never used in the construct state.

Pseudo-variation

ت ‹t› Pv ا ‹a›

Word-initially: *See* ا ‹a› Pv ت ‹t›

Word-medially: *See* ا ‹a› Pv ت ‹t›

ت ‹t› Pv أ ‹å›

Word-initially: *See* أ ‹å› Pv ت ‹t›

GRAPH ث ‹ṯ›

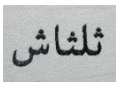
ث ‹ṯ› ~ ت ‹t›

All possible positions / Asymmetrical variation (inverse) (ث ‹ṯ› Av^{-1} ت ‹t›)
See ت ‹t› Av ث ‹ṯ›

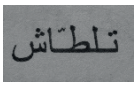
ث ‹ṯ› ~ ط ‹ṭ›

Word-medially / Restricted variation (ث ‹ṯ› Rv ط ‹ṭ›)
This variation manifests itself in graphic words in which these two graphs mark *ṭ*. Recorded only in two pairs in the corpus, shown in (129) and (130), it is marginal. The graphy ث ‹ṯ› is donor form-oriented, while ط ‹ṭ› is adoptive phonetic, based on general donor graphy.

(129)



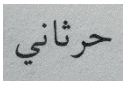
‹tlṯaš› (*Ḍa.* 99[13])
ṭəlṭaš ‘thirteen’



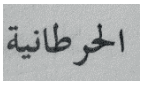
‹tlṭˑaš› (*Ṛḥ.* 95$_4$)

(cf. SA ثلاثة عشر ‹tlaṯ_ʕšr› *ṯalāṯata ʕašara* ‘idem’)

(130)



‹ḥrṯany› (*BB* 93[2])
ḥəṛṭani ‘mulatto, black’



‹[al]ḥrṭany[ħ]› (*ḤB* 8[2])
[*l-*]*ḥəṛṭani*[*ya*] ‘mulatto woman, black$^{\text{f def}}$’

(cf. SA حر ثانٍ ‹ḥr_ṯan^{in}› *ḥurr^{un} ṯanin*, lit. ‘second free man’?[122])

---

[122] The graph ث ‹ṯ› used in Form *A* in this pair can be interpreted as donor form-oriented only if the speller is assumed to have had this folk etymology: ‘a second-class free man’, ironically: ‘a slave’, in mind or associated the word in question with *ḥərrat* ‘ploughman’ (cf. SA حرّاث ‹ḥrˑaṯ› *ḥarrāṯ^{un}* ‘idem’).

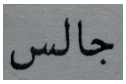
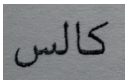
This variation, at least as exemplified in (129), is due to the fact that *ṭ* is a result of two historical phonetic processes: despirantization of *ṯ* and pharyngealization of the resulting *t*, which affected the phonetic system independently of the phonological relationships within words. Therefore, it can be hypothesized that this variation can manifest itself in all positions. Word-initially, for instance, it could be the case in a heterograph pair representing the word *ṭnaš* 'twelve'. However, only the heterograph with the phonetic graphy ط ‹ṭ›, طناش ‹ṭnaš› (*ḤB* 36_9), has been recorded, while the donor form-oriented graphy, تناش ‹ṯnaš›, is missing. Since the word-final substitability of these graphs could not be evidenced either, this variation is described as manifesting itself solely word-medially.

GRAPH ج ‹ǧ›

Before discussing the particular variants of ج ‹ǧ›, it is useful to outline the two usual functions of this graph in the MA graphy: marking *ž* and *g*. The former function, not affected by variation, is adopted from Moroccan SA. The second function, marking *g*, a sound absent from the typical SA sound inventory (a practice current in medieval Moroccan texts, Aguadé 2013: 207), is based on the principle of endogenous adaptation. It is, however, limited to cases where this sound is a result of the historical dissimilaton of reflexes of SA *ǧ* in words containing a sibilant (Heath 2002: 136-138; Aguadé 2008: 290), a process which, as it seems, cannot be observed word-finally, e.g. جلس ‹gls› *gləs* 'he sat down' (cf. SA *ǧalasa* 'idem'). Indeed, the corpus reveals that ج ‹ǧ› is never used to mark *g* which is related to *q*, e.g. *gal* 'he said' (cf. SA قال *qāla* and MA *qal* 'idem') is never spelt *جال ‹ǧal›. Hence, the principle underlying the use of ج ‹ǧ› to mark *g* is adaptive phonetic, with the adaptation being endogenous and historically motivated. In this function, ج ‹ǧ› is in variation, non-word-finally, with the three graphs discussed below.

ج ‹ǧ› ~ ك ‹k›

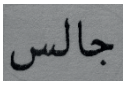
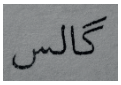
Word-initially, word-medially / Restricted variation (ج ‹ǧ› Rv ك ‹k›)
The primary function of ك ‹k› in the SA and MA graphy is to mark *k*. In MA, this graph has been adapted to mark *g* because of the phonetic similarity between these two sounds: the only difference between them being that *g* is voiced and *k* is unvoiced. The use of ك ‹k› is thus a result of phonetically motivated endogenous adaptation. Example (131) shows a word-initial occurrence of variation between ج ‹ǧ› and ك ‹k›, marginally evidenced, but also expected word-medially.

(131)  

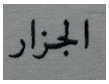
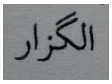
‹ğals› (*Mm.* 147_7) ‹**k**als› (*ʕD* 111^5)
galəs ‘sitting’ (participle) (cf. SA جالس ‹ğals› *ğālis*^{un} ‘idem’)

ج ‹ğ› ~ گ ‹ǧ›

Word-initially, word-medially / Restricted variation (ج ‹ğ› Rv گ ‹ǧ›)
Not being part of the SA alphabet, the graph گ ‹ǧ› is used notably in the Persian alphabet to mark *g*. The practice of using it in this function in MA is thus a result of exogenous adaptation. Examples:

(132)  

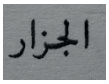
‹ğals› (*Mm.* 147_7) ‹ǧals› (*Mǧ.* 244^5)
galəs ‘sitting’ (participle) (cf. SA جالس ‹ğals› *ğālis*^{un} ‘idem’)

(133)  

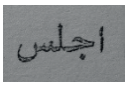
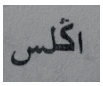
‹alğzar› (*XM* 57_6) ‹alǧzar› (*MḤ* 19^{15})
l-gəzzar ‘butcher^{def}’ (cf. SA الجزار ‹alğzar› *ʔal-ğazzār*^{u} ‘idem’)

ج ‹ğ› ~ ڜ ‹ĝ›

Word-initially, word-medially / Restricted variation (ج ‹ğ› Rv ڜ ‹ĝ›)
As indicated in Sect. III.1., graph ڜ ‹ĝ› occurs in *ḌK* and *ḤB* only. In the latter source, its connecting form is used even in the non-connecting final position, e.g. فوڜ ‹fwĝ› *fug* ‘over, upon’ (*ḤB* 16^8). This makes it impossible to determine whether the isolated form of this graph is ڜ or ڝ. Just like گ ‹ǧ›, it is not part of the SA alphabet, hence its use to mark *g* in MA is described as a result of exogenous adaptation. Variation between ج ‹ğ› and ڜ ‹ĝ› is only marginally manifested in the corpus because the use of ڜ ‹ĝ› to mark *g* being a reflex of SA *ğ* is very limited there. Examples:

(134)  

‹[al]ğzar› (*XM* 57_6) ‹ĝzar› (*ḌK* 15_{11})
[*l-*]*gəzzar* ‘butcher^{def}’ *gəzzar* ‘butcher’
(cf. SA جزار ‹ğzar› *ğazzār*^{un} ‘butcher’)

(135)  

‹aǧls› (*BT* 128[5]) ‹aĝls› (*ḤB* 10[4])

gləs ‘sit down!’ (cf. SA اجلس ‹aǧls› *ʔiǧlis* ‘idem’).

GRAPH ح ‹ḥ›

No variation.

GRAPH خ ‹x›

No variation.

GRAPH چ ‹č›

The graph چ (the *čim* of the Persian alphabet) has been used in MA graphy to mark two sounds: *g* and *č*. The former is the case if *g* corresponds diachronially to SA *ǧ* and this has been observed at least in two sources, not included in the corpus: *Tqə.* (see Aguadé 2006: 259), a literary work, and an edition of *məlḥun* poetry (al-Fāsī 1986: 25). The second function of چ ‹č›, marking *č* (Lerchundi 1900: 4 [1872: 5, footnote 1]), a sound occurring in northern and Jbala dialects (Aguadé 2008: 289-290), has not been recorded in the corpus nor has been observed by other scholars[123].

GRAPH د ‹d›

د ‹d› ~ ت ‹t›

Word-initially and word-medially / Restricted variation (د ‹d› Rv ت ‹t›)

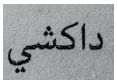
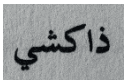
See ت ‹t› Rv د ‹d›

[123] The other manner of marking *č* indicated by Lerchundi, viz. پش , which, formally speaking, is a ش ‹š› with three additional dots below, has not been recorded in the corpus either. However, in more than one source included in it, this sound is marked, rarely and only in borrowings, by the digraph تش ‹tš›, e.g. تشامير ‹tšamyr› *čamir* ‘a type of long shirt’ (*MḤ* 60[9]), ماتش ‹matš› *mač* ‘match, game’ (*FB* 66[7]), كَايْتْشَاتِي ‹kᵃayˈtˈšᵃatʲy› *ka-yčati* ‘he is chatting (online)’ (*Mm.* 112[8]). In one isolated case, another digraph, ‹ṭš›, is used in this function: الفرناطشي ‹alfrnaṭšy› *l-fəṛnačі* ‘a person tending the water heater in public baths’ (*fəṛnatši* in Harrell 1966: 1938, from Italian *fornace* ‘furnace’). Belonging to the quantitative domain of heterography, digraphs do not concern us in this work.

د ‹d› ~ ذ ‹ḏ›

All possible positions / Asymmetrical variation (د ‹d› Av ذ ‹ḏ›)

This variation manifests itself, very frequently, in graphic words in which د ‹d› and ذ ‹ḏ› mark the sound *d*. The graphy د ‹d› is adoptive phonetic, based on general donor graphy. If the MA word has a SA cognate with *d*, as in (140) and (141), it is, in addition, coincident with the donor graphy. Heterograph pairs based on this variation mostly represent words with SA cognates which contain *ḏ*. In such cases, the graphy ذ ‹ḏ› is donor form-oriented (cf. Aguadé 2006: 257-258). Examples:

(136)  

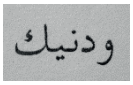
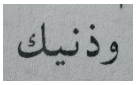
‹**d**akšy› (*ʕD* 111_9) ‹**ḏ**akšy› (*ḤB* 42^8)

dak š-ši ‘that’, lit. ‘that thing’ (cf. SA ذاك الشيء ‹ḏak_alšyo› *ḏāka š-šayʔᵘ* ‘that thing’)

(137)  

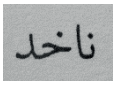
‹ġ**d**a› (*TN* 32^7) ‹ġ**ḏ**a› (*TN* 19^3)

ġədda ‘tomorrow’ (cf. SA غدا ‹ġda› *ġadan* ‘idem’)

(138)  

‹w**d**nyk› (*TM* 8_{10}) ‹w**ḏ**nyk› (*MX* 55_5)

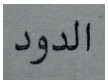
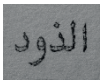
wŭdni-k ‘your ears’ (cf. SA أذنيك ‹åḏnyk› *ʔuḏunay-ka* ‘your ears$^{\text{du obl}}$’)

(139)  

‹nax**d**› (*Ṣa.* 232_8) ‹nax**ḏ**› (*BT* 116_3)

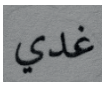
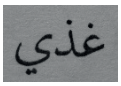
naxŭd ‘I take’ (cf. SA نأخذ ‹nåxḏ› *naʔxuḏᵘ* ‘we take’)

In less frequent, cases, ذ ‹ḏ› is used in words which have no SA cognates with corresponding *ḏ*. The reason for such spellings is analogical graphy: The speller is aware that MA *d* is spelt as ذ ‹ḏ› in many words (donor form-orientation) and uses this graphy even though the word to be written has no SA cognate with a corresponding *ḏ* that would justify this. Examples:

(140)  

‹al**d**wd› (*ʕD* 79_{10}) ‹al**ḏ**wd› (*BT* $_3611$)

d-dud ‘worms$^{\text{def}}$’ (cf. SA الدود ‹aldwd› *ʔad-dūdᵘ* ‘idem’)

(141)  

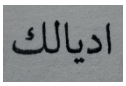
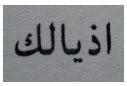
‹ġ**d**y› (*LŠ* 54[9]) ‹ġ**ḏ**y› (*Mn.* 23[6])

ġadi ‘going to’ (future marker) (cf. SA غاد ‹ġad› *ġādin* ‘going; becoming’).

(142)  

‹**d**aba› (*BB* 97_{10}) ‹**ḏ**aba› (*MX* 203[14])

daba ‘now’

(143)  

‹a**d**yalk› (*FL* 111_{10}) ‹a**ḏ**yalk› (*XM* 113[8])

dyal-ək ‘yours, of yours’

د ‹d› ~ ض ‹ḍ›

All possible positions / Restricted variation (د ‹d› Rv ض ‹ḍ›)

This variation manifests itself first and foremost in graphic words in which these two graphs mark a pharyngealized *ḍ*. In marginal cases, the sound marked by them is plain (non-pharyngealized) *d*. The variation is related to the fact that in numerous cases, plain sounds in SA phonetic words have pharyngealized correspondents in MA cognates (cf. e.g. Aguadé 2008: 290) (and, in a few examples, *vice versa*). If the sound marked by these graphs is *ḍ*, the corresponding sound in the SA cognate is usually *d* or *ḍ*. Apart from these two major situations, a third possibility is that a SA cognate has *ḏ* or that there is no SA cognate at all.

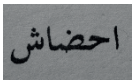
In the first situation (*ḍ* in MA, *d* in the SA cognate), the graphy د ‹d› is donor form-oriented[124], while ض ‹ḍ› is adoptive phonetic, based on general donor graphy. Examples:

(144)  

‹**d**wrħ› (*FʕḤ* 4[2]) ‹**ḍ**wrħ› (*ṚḥḤ* 156[2])

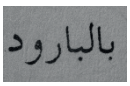
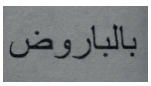
ḍuṛa ‘walk, tour; turn’ (cf. SA دورة ‹dwrħ› *dawra*^{tun} ‘turn, rotation’)

[124] As observed by Aguadé (2006: 258, 4.3 and 2013: 210, 214), in the case of some words, ‘the weight of the orthography of the classical language’ prevents the speller from using the phonetic spelling ض ‹ḍ› (2013: 214). The analysis of the corpus confirms this: Donor form-oriented graphy د ‹d› is used consistently for writing such words as: *ḍaṛ* ‘house’, *ṣəḥḍ* ‘heat’, *ṛəḍḍ* ‘he returned’ and others.

(145)  

‹ḥ˙**d**aš˙› (*ṮM* 7_6) ‹(a)ḥ**ḍ**aš› (*DB* 212^7)

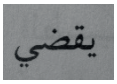
ḥḍaš ‘eleven’ (cf. SA أحد عشر ‹åḥd_ʕšr› *ʔaḥada ʕašara* ‘idem’)

(146)  

‹balbarw**d**› (*Fʕ* 56^5) ‹balbarw**ḍ**› (*Ṛḥ.* 175^6)

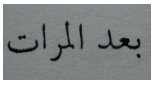
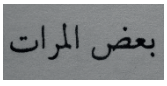
bə-l-ḥaṛuḍ ‘with gunpowder^{def}’ (cf. SA بالبارود ‹balbarwd› *bi-l-bārūd*^{i} ‘idem’)

The second situation (*ḍ* in both MA and the SA cognate) is illustrated in (147) and (148).

(147)  

‹yq**d**y› (*ḌK* 24_4) ‹yq**ḍ**y› (*MX* 14^3)

yəqḍi ‘he accomplishes’ (cf. SA يقضي ‹yqḍy› *yaqḍī* ‘idem’)

(148)  

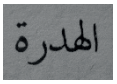
‹bʕ**d**_almrat› (*Ṣa.* 223^{11}) ‹bʕ**ḍ**_almrat› (*Ṣa.* 241^5)

bəʕḍ əl-məṛṛat ‘sometimes’ (cf. SA بعض المرات ‹bʕḍ_almrat› *baʕḍ*^{a} *l-marrāt*^{i} ‘idem’)

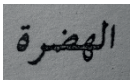
In this situation, the graphy ض ‹ḍ› is adoptive phonetic, based on general donor graphy, and is coincident with the donor graphy, while د ‹d› can be interpreted as analogical graphy: The speller is aware that MA *ḍ* is sometimes spelt د ‹d› (donor form-orienation), and (s)he uses this graphy even though the word to be written does not have a SA cognate with *d* that would justify this. Analogical graphy may be suggested to spellers by the fact that in MA there are pairs of words in which *d* and *ḍ* are mutually substitutable, e.g. *dəbbəṛ* ~ *ḍəḅḅəṛ* ‘he arranged’, *hdər* ~ *hḍəṛ* ‘he talked’, which is often reflected in writing, resulting in the parallelism of double pronunciation and double graphy (cf. a similar case of س ‹s› and ص ‹ṣ› below).

In the third situation, there is no SA cognate containing *d* or *ḍ*. In (149), the SA cognate has *ḏ*, while (150) shows a word with no SA cognate at all.

(149)



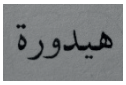
‹alh**d**rħ› (*Mn.* 78$_4$)



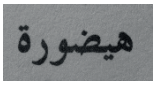
‹alh**ḍ**rħ› (*Mi.* 44$_{10}$)

l-həḍṛa ‘conversation, talk^{def}’ (cf. SA الهذر ‹alhḏr› *ʔal-haḏr*^{u} ‘prattle^{def}’)

(150)



‹hy**d**wrħ› (*Mm.* 152$_1$)



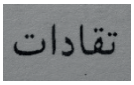
‹hy**ḍ**wrħ› (*Mu.* 29$_{14}$)

hiḍuṛa ‘(sheep)skin’

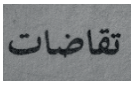
In these cases, the graphy ض ‹ḍ› is adoptive phonetic, based on general donor graphy, while د ‹d› is analogical (cf. explanation above).

Finally, this variation is marginally evidenced in graphic words representing words in which the two graphs mark *d*, while the SA cognate has *ḍ*. The only heterograph pair showing this which has been recorded in the corpus is given in (151).

(151)



‹tqa**d**at› (*ḌK* 15$_{10}$)



‹tqa**ḍ**at› (*ḤB* 14$_3$)

tqadat ‘she/it^{f} ended’ (cf. SA تقضّت ‹tqaḍt› *taqaḍḍat* ‘idem’)

The graphy د ‹d› is adoptive phonetic, based on general donor graphy, while ض ‹ḍ› is donor form-oriented.

GRAPH ذ ‹ḏ›

ذ ‹ḏ› ~ د ‹d›

All possible positions / Asymmetrical variation (inverse) (ذ ‹ḏ› Av^{-1} د ‹d›)

See د ‹d› Av ذ ‹ḏ›

ذ ‹ḏ› ~ ض ‹ḍ›

{All possible positions / Restricted variation (ذ ‹ḏ› Rv ض ‹ḍ›)}

Restricted variation between ذ ‹ḏ› and ض ‹ḍ› is expected to manifest itself in two cases: First, in graphic words representing words containing *ḍ* which is a result of the despirantization and pharyngealization of *ḏ*, e.g. *fxaḍ* ‘thighs’, for which phonetic graphy, فخاض ‹fxaḍ› (*Fʕ* 54$_8$), has been recorded in the corpus, while the donor-

oriented form (فخاذ ‹fxaḏ›) is expected. Second, this variation is expected to manifest itself in graphic words representing words containing *ḍ* and having no SA cognates, e.g. *ḍʕiṛa* 'a fine', for which an analogical graphy with ذ ‹ḏ›, ذعيرة ‹ḏʕyrħ› (*MX* 41$_7$), has been recorded, while its phonetic graphy (ضعيرة ‹ḍʕyrħ›) is expected.

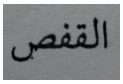
GRAPH ر ‹r›

No variation.

GRAPH ز ‹z›

ز ‹z› ~ ص ‹ṣ›

Word-finally / Restricted variation (ز ‹z› Rv ص ‹ṣ›)
Although this variation manifests itself in one pair only, both graphic words are frequent enough to justify taking it into consideration. The graphs bound by it mark the sound *z*. The graphy ز ‹z› in this function is adoptive phonetic, based on general donor graphy, while ص ‹ṣ› results from donor form-orientation[125].

(152)  

‹alqf**z**› (*Ṣa.* 246[8]) ‹alqf**ṣ**› (*Ṣa.* 247[6])
lə-qfəz 'cage[def]' (cf. SA القفص ‹alqfṣ› *ʔal-qafaṣ*^{u} 'idem')

GRAPH س ‹s›

س ‹s› ~ ص ‹ṣ›

All possible positions / Restricted variation (س ‹s› Rv ص ‹ṣ›)
As was the case with *d* and *ḍ* (cf. "د ‹d› Rv ض ‹ḍ›"), in many words, the SA plain (non-pharyngealized) *s* corresponds to the pharyngealized *ṣ* in MA. Predominantly, variation between س ‹s› and ص ‹ṣ› manifests itself in graphic words represent-

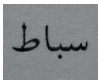
[125] It is interesting to note that ز ‹z› is used in MA to denote pharyngealized *ẓ* (IPA: [zˤ]) (e.g. in زرابي ‹zraby› *ẓṛabi* 'carpets, rugs') although the SA graph ظ ‹ẓ› could fulfil this function better from the point of view of sound-letter-correspondence because it represents this sound in some regional pronunciations of SA, e.g. Egyptian (Cairene) (Holes 2004: 59). However, in Moroccan SA, it is usually pronounced *ḍ* (IPA: [dˤ]) and hence it is not used in this function. The authors of works included in the corpus use ظ ‹ẓ› only in some donor form-oriented or analogical graphies.

ing phonetic words that contain a pharyngealized *ṣ*, less frequently a plain *s*. It has been recorded word-initially and word-medially and is also expected to occur word-finally.

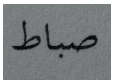
If the sound marked by these graphs is *ṣ*, the corresponding sound in the SA cognate is usually *s* or *ṣ*. A third possibility is that the word has no SA cognate which could be taken into consideration as a graphic point of reference.

In the first situation (*ṣ* in MA, *s* in the SA cognate), the graphy س ‹s› is donor form-oriented, while ص ‹ṣ› is adoptive phonetic, based on general donor graphy. Examples:

(153)

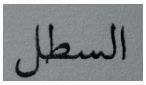


‹**s**baṭ› (*ḌK* 17[2])
ṣəbḅaṭ ‘shoe’ (cf. SA سبّاط ‹sbˑaṭ› *sabbāṭ*[un] ‘idem’)

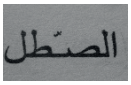


‹**ṣ**baṭ› (*Fʕ* 15$_3$)

(154)



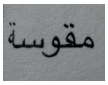
‹al**s**ṭl› (*MḤ* 64[4])
ṣ-ṣṭəḷ ‘bucket[def]’ (cf. SA السطل ‹alsṭl› *ʔas-saṭl*[u] ‘idem’)



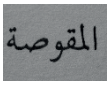
‹al**ṣ**ˑṭl› (*Ṛḥ.* 153[5])

The second situation (*ṣ* in both MA and the SA cognate) is exemplified in one heterograph pair only:

(155)



‹mqw**s**ħ› (*ṮN* 10[13])
mquṣa ‘scissors[pl]’
(cf. SA مقصّ ‹mqṣˑ› *miqaṣṣ*[un] ‘scissors’)



‹[al]mqw**ṣ**ħ› (*Ṣa.* 226$_{12}$)
[*lə-*]*mquṣa* ‘scissors[pl def]’

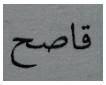
Here, the graphy ص ‹ṣ› is adoptive phonetic, based on general donor graphy and coincident with the donor graphy, while س ‹s› is an instance of analogical graphy: The speller is aware that the MA sound *ṣ* is marked as س ‹s› in some words (donor form-orientation) and uses this graphy even though the word to be written does not have a SA cognate with *s* that would justify this graphy. Analogical graphy may be suggested to spellers by the fact that in MA there are pairs of words in which *s* and *ṣ* are mutually substitutable, e.g. *sag* ~ *ṣag* ‘leg’, *safəṛ* ~ *ṣafəṛ* ‘he travelled’, which is often reflected in writing, resulting in the parallelism of double pronunciation and double graphy (cf. a similar case of د ‹d› and ض ‹ḍ› discussed above). It is, however, also possible that the graphy س ‹s› in (155) is meant

to reflect depharyngealized pronunciation in this word, recorded by Heath (2002: 440) as *mqəss*.

In the third situation, the MA word containing *ṣ* has no transparent SA cognate that could justify donor form-oriented graphy. Examples:

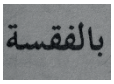
(156)

‹qa**s**ḥ› (*ḌḎ* 212[7])
qaṣəḥ ‘hard’ (cf. SA قاسح *qāsiḥ*[un] ‘idem’, very rare)

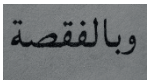


‹qa**ṣ**ḥ› (*ṮM* 51$_3$)

(157)



‹balfq**s**ħ› (*ḤB* 52[7])
bə-l-fəqṣa
‘out of sorrow[def]’
(no transparent SA cognate)



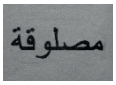
‹[w]balfq**ṣ**ħ› (*MX* 164$_1$)
[*u*-]*bə-l-fəqṣa*
‘[and] out of sorrow[def]’

In such cases, the graphy with ص ‹ṣ› is adoptive phonetic, based on general donor graphy, while س ‹s› is analogical, as explained under (155).

Finally, in a rather rare configuration of this variation, the sound marked by س ‹s› and ص ‹ṣ› is *s*. Examples:

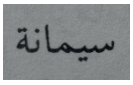
(158)

‹[al]m**s**lwq› (*ḤM* 83$_3$)
[*l*-]*məsluq* ‘boiled[def]’
(cf. SA مسلوق ‹mslwq› *maslūq*[un] ‘boiled’)

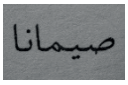


‹m**ṣ**lwq[ħ]› (*ṚḤ.* 199[9])
məsluq[*a*] ‘boiled[f]’

(159)



‹**s**yman(ħ)› (*FʕḎ* 58[13])
simana ‘week’



‹**ṣ**yman(a)› (*ṮM* 38$_5$)

The graphy س ‹s› is adoptive phonetic, based on general donor graphy – and in the case of (158), which has a SA cognate, coincident with donor graphy. The graphy ص ‹ṣ› might seem to be analogical, but the usage evidenced in the corpus does not make it possible to identify its exact basis because the sources analyzed do not reveal instances of MA *s* marked as ص ‹ṣ› by virtue of one of the proposed spelling principles. In particular, no MA words with *s* spelt as ص ‹ṣ› due to a possible influence of SA graphy have been recorded.

GRAPH ش ‹š›

No variation.

GRAPH ص ‹ṣ›

ص ‹ṣ› ~ س ‹s›

All possible positions / Restricted variation (ص ‹ṣ› Rv س ‹s›)
See س ‹s› Rv ص ‹ṣ›

ص ‹ṣ› ~ ز ‹z›

Word-finally / Restricted variation (ص ‹ṣ› Rv ز ‹z›)
See ز ‹z› Rv ص ‹ṣ›

GRAPH ض ‹ḍ›

ض ‹ḍ› ~ د ‹d›

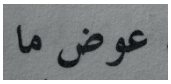
All possible positions / Restricted variation (ض ‹ḍ› Rv د ‹d›)
See د ‹d› Rv ض ‹ḍ›

ض ‹ḍ› ~ ذ ‹ḏ›

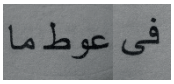
{All possible positions / Restricted variation (ض ‹ḍ› Rv ذ ‹ḏ›)}
See {ذ ‹ḏ› Rv ض ‹ḍ›}

ض ‹ḍ› ~ ط ‹ṭ›

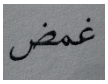
Word-finally / Restricted variation (ض ‹ḍ› Rv ط ‹ṭ›)
This variation manifests itself, rather marginally, in graphic words in which these graphs mark the sound *ṭ* resulting from the devoicing of *ḍ*. The graphy ط ‹ṭ› is adoptive phonetic, based on general donor graphy, while ض ‹ḍ› represents *ḍ*, the last radical of the lexical root, morphologically. Note that this case is different from the majority of morphological graphies, which are used to mark affixes, not parts of the root. Examples:

(160) 

‹ʕwḍ_ma› (*Mu.* 15_5)
ʕiwəḍ [ʕɨwətˤ] *ma*
(cf. SA عوض ‹ʕwḍ› *ʕiwaḍ*^{a} ‘idem’)



‹[fá_]ʕwṭ_ma› (*Ṛḥ.* $186_{8\text{-}7}$)
[*f*-]*ʕiwəḍ* [fʕɨwətˤ] *ma* ‘instead of’

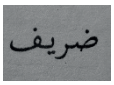
(161) 

‹ġmḍ› (*Mn.* 34_1)
ġəmməḍ [ɣəmːətˤ] ‘close!’ (said of eyes)
(cf. SA غمّض ‹ġmːḍ› *ġammiḍ* ‘idem’)

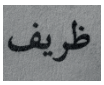
‹ġmːṭ› (*WM* 31_6)

ض ‹ḍ› ~ ظ ‹ẓ›

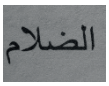
All possible positions / Asymmetrical variation (ض ‹ḍ› Av ظ ‹ẓ›)
This variation usually manifests itself in graphic words which have SA cognates with a corresponding ظ ‹ẓ›. Both variants mark *ḍ*. The graphy ض ‹ḍ› is adoptive phonetic, based on general donor graphy, while ظ ‹ẓ› is donor form-oriented. Examples:

(162) 

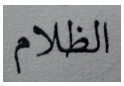
‹ḍryf› (*ṬM* 48^6)
ḍṛīf ‘nice’ (cf. SA ظريف ‹ẓryf› *ẓarīf*^{un} ‘idem’)



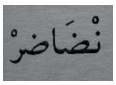
‹ẓryf› (*Mu.* 16_3)

(163) 

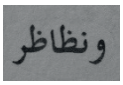
‹alḍlam› (*Ṛḥ.* 139_1)
ḍ-ḍlam ‘darkness^{def}’ (cf. SA الظلام ‹alẓlam› *ʔaẓ-ẓalām*^{u} ‘idem’)



‹alẓlam› (*MḤ* 22_1)

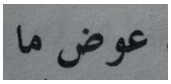
(164) 

‹nˑḍᵃaḍrˑ› (*Mm.* 119_3)
nḍaḍəṛ ‘glasses’
(cf. SA نظارة ‹nẓarħ› *naẓẓāra*^{tun} ‘idem’)

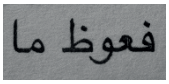


‹[w]nẓaẓr› (*ḤB* 13^7)
[*u*-]*nḍaḍəṛ* ‘[and] glasses’

Marginally, this variation manifests itself in graphic words with a SA cognate spelt with ض ‹ḍ›, as in (165).

(165) 

‹ʕwḏ_ma› (*Mu.* 15_5)
ʕiwəḍ ma
(cf. SA عوض ‹ʕwḍ› *ʕiwaḍ*[a] ‘idem’)



‹[f]ʕwẓ_ma› (*TN* 12^6)
[*f*-]*ʕiwəḍ ma* ‘instead of’

In this case, the graphy ظ ‹ẓ› is analogical: The speller is aware that the MA sound *ḍ* is sometimes written ظ ‹ẓ› (donor form-orientation) and uses this graphy even though the graphy of the SA cognate does not justify this.

GRAPH ط ‹ṭ›

ط ‹ṭ› ~ ت ‹t›

All possible positions / Restricted variation (ط ‹ṭ› Rv ت ‹t›)
See ت ‹t› Rv ط ‹ṭ›

ط ‹ṭ› ~ ث ‹ṯ›

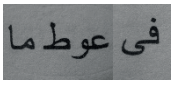
Word-medially / Restricted variation (ط ‹ṭ› Rv ث ‹ṯ›)
See ث ‹ṯ› Rv ط ‹ṭ›

ط ‹ṭ› ~ ض ‹ḍ›

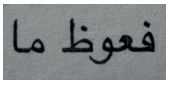
Word-finally / Restricted variation (ط ‹ṭ› Rv ض ‹ḍ›)
See ض ‹ḍ› Rv ط ‹ṭ›

ط ‹ṭ› ~ ظ ‹ẓ›

Word-finally / Restricted variation (ط ‹ṭ› Rv ظ ‹ẓ›)
This variation is marginal as it has manifested itself in only one heterograph pair. The graphs bound by it mark the sound *ṯ* resulting from the devoicing of *ḍ*.

(166) 

‹f(á_)ʕwṯ_ma› (*Ṛḥ.* $186_{8\text{-}7}$)
f-ʕiwəḍ [fʕɨwətˁ] *ma* ‘instead of’
(cf. SA عوض ‹ʕwḍ› *ʕiwaḍ*[a] ‘idem’)



‹fʕwẓ_ma› (*TN* 12_6)

The graphy ط ‹ṭ› is adoptive phonetic, based on general donor graphy, while ظ ‹ẓ› is an instance of analogical graphy (explained under "ض ‹ḍ› Av ظ ‹ẓ›").

GRAPH ظ ‹ẓ›

ظ ‹ẓ› ~ ض ‹ḍ›

All possible positions / Asymmetrical relation (inverse) (ظ ‹ẓ› Av^{-1} ض ‹ḍ›)
See ض ‹ḍ› Rv ظ ‹ẓ›

ظ ‹ẓ› ~ ط ‹ṭ›

Word-finally / Restricted variation (ظ ‹ẓ› Rv ط ‹ṭ›)
See ط ‹ṭ› Rv ظ ‹ẓ›

GRAPH ع ‹ʕ›

No variation.

GRAPH غ ‹ġ›

The primary function of the graph غ ‹ġ› is marking *ġ* (IPA [ɣ]), voiced velar fricative, in both SA (Beeston 1970: 18; Holes 2004: 58) and MA (Kjamilev 1968: 19; Aguadé 2003: 85; Caubet 2008: 275)[126]. In MA graphy, it has been adapted to mark *g*, the voiced velar plosive sound (Harrell 1962: 3, Aguadé 2003: 83)[127]. The reason for this adaptation is that the sound it represents in the donor language shares some phonetic features with the sound for the spelling of which it was adapted: voicedness and velar articulation. This is thus a case of phonetically motivated endogenous adaptation.

---

[126] Other points of articulation have also been given for this sound. In SA, it has been characterized as: uvular (e.g. al-Ani 2008: 598), postvelar (e.g. Corriente 2002: 24) or context-conditioned: postdorsal-postvelar in the vicinity of *i* or postdorsal-uvular in the vicinity of *a* and *u* (Käshtner 1981: 65-66). In MA, it has also been described as uvular in Harrell (1962: 3) and as prevelar in Durand (2004: 70).

[127] Kjamilev (1968: 22) and Caubet (2008: 275) characterize its point of articulation as postpalatal.

The use of غ ‹ġ› in the function of marking *g* is, however, marginal and the scarcity of graphic words which show this use makes it difficult to compose heterograph pairs. This graph occurs in this function in only three graphic words in the corpus: غارو *garṛu* 'a cigaret' (*ṮM* 11^{10}), أغادير *agadir* 'Agadir' (*XM* 146_{7}), and هامبرغر *hambŭrgər* 'hamburger' (*XF* 96_{6}), none of which is of Arabic origin. These graphic words evidence only word-initial and word-medial use of غ ‹ġ› marking *g*. Word-final use can be expected, although it would be very marginal too.

غ ‹ġ› ~ ك ‹k›

Word-initially, word-medially / Restricted variation (غ ‹ġ› Rv ك ‹k›)
Both graphs bound by this marginal variation mark *g* as a result of phonetically motivated endogenous adaptation (see "ج ‹ğ› Rv ك ‹k›" for the discussion of ك ‹k›). The use of غ ‹ġ› in this function is marginal, while that of ك ‹k› very frequent. Example:

(167)

‹**ġ**arw› (*ṮM* 11^{10})
garṛu 'cigaret'

‹[l]**k**arw› (*ḤM* 89^{10})
[*l-*]*garṛu* 'cigaret$^{\text{def}}$'

غ ‹ġ› ~ گ ‹ǵ›

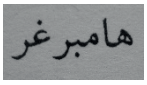
Word-initially, word-medially / Restricted variation (غ ‹ġ› Rv گ ‹ǵ›)
While the principle underlying the marginal use of غ ‹ġ› to mark *g* is that of phonetically motivated endogenous adaptation, the frequent graphy گ ‹ǵ› in this function is based on exogenous adaptation. Examples:

(168)

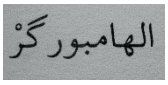
‹**ġ**arw› (*ṮM* 11^{10})
garṛu 'cigaret'

‹**ǵ**arw› (*MḤ* 16^{11})

(169)



‹hambr**ġ**r› (*XF* 96_{6})
hambŭrgər
'hamburger'



‹[al]hamb(w)r**ǵ**r› (*ṮM* 65^{1})
[*l-*]*hambŭrgər*
'hamburger$^{\text{def}}$'

غ ‹ġ› ~ ڠ ‹ĝ›

{Word-initially, word-medially / Restricted variation (غ ‹ġ› ~ ڠ ‹ĝ›)}
This variation has not manifested itself in the corpus but is expected because ڠ ‹ĝ› is a free variant of گ ‹ǵ›, which, in its turn, is a variant of غ ‹ġ› (see above). The graphic words ڠارو ‹ĝarw› for *garṛru* 'cigaret' and هامبرڠر ‹hambrĝr› for *hambŭrgər* 'hamburger' are expected heterographs of those given under "غ ‹ġ› Rv گ ‹ǵ›".

GRAPH ڤ ‹f̣›

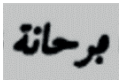
ڤ ‹f̣› ~ ف ‹f›

All possible positions / Restricted variation (ڤ ‹f̣› Rv ف ‹f›)
This variation manifests itself in graphic words in which these graphs mark *f*. Both graphies are adoptive phonetic, with ڤ ‹f̣› being based on local donor graphy and ف ‹f› being based on general donor graphy. In this function, the shape of the latter in non-connecting positions is ف ‹f_f›.

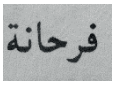
This variation is restricted for the following reasons: On the one hand, the graph ڤ ‹f̣› cannot be substituted for ف ‹f› if the latter marks *q*. On the other hand, ف ‹f› cannot be substituted for ڤ ‹f̣› in graphic words or texts which contain ف ‹f› marking *q* (e.g. in فڤة ‹ff̣ẖ› *qəffa* 'basket' to yield *ففة ‹ffẖ›) because this would result in ف ‹f› marking two different sounds according to two different writing traditions in a single graphic word or text[128].

The variation is marginal because the analyzed pages of the only source in the corpus in which the local Maghrebi graphy is, very modestly, used, viz. *ḌK*, yielded only two occurrences of ڤ ‹f̣›: It was used twice in one graphic word, written decoratively on the opening page of a chapter, while in the body of the text, ف ‹f› is used, as shown in (171). Form *A* in (170) is taken from a source not included in the corpus.

(170)

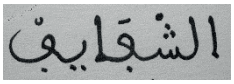
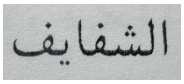


‹f̣rḥanẖ› (*Ḥaš.* 36)
fəṛḥana 'happy[f]'



‹frḥanẖ› (*MX* 71_1)

[128] In other words, ف ‹f› marking *f* is a conditioned graphy, i.e. when substituting this graph for ڤ ‹f̣› (which always marks *f*), any ف ‹f› used originally to mark *q* must be substituted for by ق ‹q›. By contrast, the graphy ڤ ‹f̣› is unconditioned, i.e. its use does not require any other graph – in particular, ق ‹q› marking *q* – to be substituted for by any other graphs, and consequently ڤ ‹f̣› and ق ‹q› can be used in a single text (as is the case in *Ḥaš.*).

(171)  

‹alṣ̌**f̣**^{a}ay**f̣**› (*ḌK* 20$_2$) ‹alš**f**ay**f**$_{f}$› (*ḌK* 22^4)

š-šfayəf ‘lips^{def}’

As expected, the *rasm* of ف ‹f› in the non-neutralized (non-connecting) position in (171) is the shallow-bowled (fricative) one: ف ‹f$_f$›.

GRAPH ف ‹f›

ف ‹f› ~ ب ‹f̣›

All possible positions / Restricted variation (‹f› Rv ب ‹f̣›)

See ب ‹f̣› Rv ف ‹f›

ف ‹f› ~ ق ‹q›

All possible positions / Restricted variation (ف ‹f› Rv ق ‹q›)

This variation manifests itself in graphic words in which the two graphs mark *q*. It has been recorded word-medially, while word-initially and word-finally it is an expected variation. Both graphies are adoptive phonetic, with ف ‹f› being based on local donor graphy and ق ‹q› being based on general donor graphy. The *rasm* of ف ‹f› in non-connecting positions is expected to be the deep-bowled (plosive) one: ف ‹f$_q$›. This, however, could not be verified on the basis of the corpus.

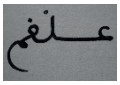
The variation is restricted because it manifests itself only if the graphs mark the sound *q*, which is not always the case, because both are phonetically ambiguous. In addition, ف ‹f› marking *q* cannot be used in graphic words or texts which contain ف ‹f› marking *f* (e.g. in قفة ‹qfh̄› *qəffa* ‘basket’ to yield *ففة ‹ffh̄›) because this would result in ف ‹f› marking two different sounds according to two different writing traditions in a single graphic word or text[129].

As was the case with the preceding variation, this one is marginal too: the analyzed pages of *ḌK*, the only source in the corpus in which the local Maghrebi

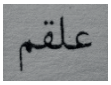
[129] In other words, the graphy ف ‹f› marking *q* is conditioned, i.e. its use in a graphic word or text requires any ف ‹f› used there to mark *f* to be substituted for by ب ‹f̣›, in order to avoid ف ‹f› marking two different sounds according to two different writing traditions in a single graphic word or text. By contrast, the graphy ق ‹q› marking *q* is unconditioned, i.e. its use does not require any other graphs – in particular, ب ‹f̣› – to be substituted for by any other graphs and consequently ق ‹q› and ب ‹f̣› can be used in a single text (as is the case in *Ḥaš.*).

graphy is used, yielded only one occurrence of ڢ ‹f› marking *q*, used, word-medially[130], in a graphic word written decoratively on a chapter-opening page. In the body of text, the general donor-graphy, ق ‹q›, is used:

(172)



‹ʕlˑ**f**m› (*ḌK* 7$_1$)
ʕəlqəm ‘colocynth’



‹ʕl**q**m› (*ḌK* 8^7)

ڢ ‹f› ~ ڤ ‹v›

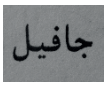
Word-initially, word-medially / Restricted variation (ڢ ‹f› Rv ڤ ‹v›)
This variation manifests itself in graphic words in which these graphs mark the sound *v*. As a rule, such graphic words represent loanwords. The use of ڢ ‹f› in this function is based on phonetically conditioned endogenous adaptation: In the donor language, this graph marks *f*, i.e. the unvoiced counterpart of *v*. As for ڤ ‹v›, it does not form part of the traditional SA alphabet[131] and its use to mark *v* in MA is based on exogenous adaptation[132]. Examples:

(173)

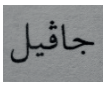
‹**f**yla› (*Mi.* 61$_8$)
villa ‘villa’

‹**v**yla› (*TN* 21$_5$)

(174)



‹ğa**f**yl› (*XM* 148^7)
žavil ‘household bleach’[133]



‹ğa**v**yl› (*TM* 9$_3$)

---

[130] Therefore, the shape of its non-connecting *rasm* cannot be ascertained.

[131] It is sometimes used in SA texts for marking *v* in borrowings, especially foreign proper names.

[132] Both graphs are conditioned graphies. The use of ڢ ‹f› to mark *v* in a graphic word or text requires any ڢ ‹f› used there to mark *q* to be substituted for by ق ‹q›, in order to avoid ڢ ‹f› marking two different sounds according to two different writing traditions in a single graphic word or text. The use of ڤ ‹v› to mark *v* in a graphic word or text requires any ڤ ‹v› used there to mark *g* to be substituted for by some other graph: ق ‹q›, ك ‹k›, گ ‹ǵ›, ڴ ‹ĝ› or غ ‹ġ› in order to avoid ڤ ‹v› marking ambiguously *v* and *g* in a single graphic word or text.

[133] The word *žavil* derives from the French *eau de Javel*, lit. ‘Javel water’, a disinfecting and bleaching chemical solution for domestic use, named after a district in Paris where it was originally produced.

It is rather unlikely for this variation to manifest itself word-finally because in order for this to occur, these graphs would have to mark a word-final *v*, which is not allowed by the phonology of MA (see, however, under ڤ ‹v› for this graph used to mark a word-final *v* in an Arabic transcription of a French word).

GRAPH ق ‹q›

ق ‹q› ~ ف ‹f›

All possible positions / Restricted variation (ق ‹q› Rv ف ‹f›)

See ف ‹f› Rv ق ‹q›

ق ‹q› ~ ڤ ‹v›

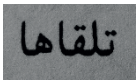
All possible positions / Restricted variation (ق ‹q› Rv ڤ ‹v›)

This variation manifests itself in graphic words in which these graphs are used to mark *g*, a pronunciation ascertained for ق ‹q› in (175) and (177) and assumed in (176).

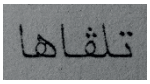
(175) 

‹**q**aʕ› (*Ṣa.* 233_{10})
gaʕ ‘even’, intensifying particle

‹**v**aʕ› (*FB* 65^{11})

(176) 

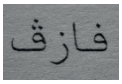
‹tl**q**aha› (*XM* 149_{4})
təlga-ha ‘you find it^{f}’ (cf. SA تلقاها ‹tlqaha› *talqā-hā* ‘idem’)



‹tl**v**aha› (*FB* 70_{7})

(177) 

‹faz**q**› (*Mn.* 29_{12})
fazəg ‘wet’



‹faz**v**$_{f}$› (*FB* 6_{2})

For graphic words with SA cognates with *q*, such as (176), the use of ق ‹q› in this function is explainable by two related principles. First, it can be interpreted as donor form-oriented. The other possible principle is that of historically mo-

tivated endogenous adaptation: ق ‹q› is adapted to mark *g* because this sound is a reflex of SA and MA *q*. For instance, *g* in the MA *təlga-ha* ‘you find it[f]’ corresponds to SA *q* in *talqā-hā* ‘idem’ and to MA *q* in *təlqa-ha* ‘idem’. If there is no SA cognate, as in (175) and (177), its use is an instance of analogical graphy: The speller is aware that ق ‹q› is used to mark *g* in a number of words (donor form-orientation) and uses this graphy in a word lacking a SA cognate that could justify this graphy. As for ڤ ‹v›, a graph which is not part of the SA alphabet, its use is a result of exogenous adaptation (apart from this function, it also marks *v*, see “ف ‹f› Rv ڤ ‹v›”). It is used in this function, quite regularly, only in one source, *FB* (with one isolated occurrence recorded in *MX*). There, *g* marked as ڤ ‹v› is always a reflex of SA *q*, never *ǧ*[134], but it may also occur in words with no SA cognates.

Contrary to expectations based on the considerations presented in Sect. III.1., the unconnected shape of ڤ ‹v› used word-finally in *FB* – illustrated in (177) – is not ڧ ‹v_q›, i.e. the deep-bowled plosive *rasm*, as the historical relationship between the two plosives *q* and *g* would suggest, but ڣ ‹v_f›, i.e. the shallow-bowled fricative *rasm*, used in graphs typically marking *f* or *v*.

ق ‹q› ~ ك ‹k›

All possible positions / Restricted variation (ق ‹q› Rv ك ‹k›)

This variation manifests itself in graphic words in which both graphs are used to mark *g* being a reflex of SA *q*. The graph ق ‹q› used in most graphic words which could serve as heterographs based on this variation is phonetically indeterminate (see Sect. IV.2.1.)[135]. In (178), the pronunciation *g* seems to be ascertained because elsewhere in this source this word as said by the same character is written with the univocal graphy ڨ ‹ĝ›. In order to ascertain that the heterographs with ق ‹q› and ك ‹k› in (179) are homophonous, this pair has been taken from a single source in which each heterograph represents the same word uttered by a single character of the same narrative. The homophony of the heterographs in (180) is assumed on the basis that both come from works of a single author.

[134] Al-Fāsī (1986: 25) adopts this convention explicitly.

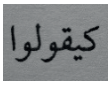
[135] The graph ‹k›, too, is phonetically ambiguous, but not phonetically indeterminate, i.e. its phonetic function can be determined by resorting to the context. The corpus, however, reveals a certain reluctancy to use it in words which would make this necessary. For inśtance, only exceptionally are *gəlb* ‘heart’ or *gal* ‘he said’ written كلب ‹klb› and كال ‹kal›, respectively, because they can be misśtakenly pronounced as *kəlb* ‘dog’ and *kal* ‘he ate’, especially when taken out of context.

(178)

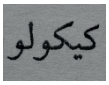
‹**q**al› (*ḤB* 4[4])
gal ‘he said’ (cf. SA قال ‹qal› *qāla* ‘idem’)

‹**k**al› (*ʕD* 202$_{10}$)

(179)

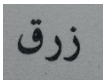


‹ky**q**wlw(a)› (*ṢX* 95$_{10}$)
ka-ygulu ‘they say’ (cf. SA يقولوا ‹yqwlwa› *yaqūlū* ‘that they say’)



‹ky**k**wlw› (*ṢX* 95$_{6}$)

(180)



‹zr**q**› (*Mu.* 57[3])
ẓṛəg ‘blue’ (cf. SA أزرق ‹åzrq› *ʔazraq*^{u} ‘idem’)

‹zr**k**› (*ḤB* 16$_{3}$)

Marking *g* is not a function of either of these two graphs in Moroccan SA, from which this sound is absent, nor their primary function in MA, in which ق ‹q› normally marks *q* and ك ‹k› normally marks *k*. Here, the use of ك ‹k› to mark *g*, its voiced counterpart, is a result of phonetically motivated endogenous adaptation. As for the use of ق ‹q› in this function, it is based either on the principle of donor form-orientation or that of historically motivated endogenous adaptation (see “ق ‹q› Rv ڤ ‹v›).

To a considerably lesser extent this variation involves ق ‹q› marking *g* in graphic words which have no SA cognate with a corresponding *q*. In this case, the graphy is analogical (see, again, “ق ‹q› Rv ڤ ‹v›). Examples:

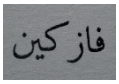
(181)

‹**q**aʕ› (*Ṣa.* 233$_{10}$)
gaʕ ‘indeed, at all’

‹**k**aʕ› (*Ṣa.* 214$_{2}$)

(182)

‹faz**q**› (*Mn.* 29$_{12}$)
fazəg ‘wet’



‹faz**k**[yn]› (*Mn.* 77[9])
fazg[*in*] ‘wet[pl]’

All possible positions / Restricted variation (ق ‹q› Rv گ ‹ǵ›)
This variation is similar to that between ق ‹q› and ك ‹k› in that the sound marked by the graphs is *g*. In (183) and (185), the graphy ق ‹q› is phonetically indeterminate. For (183), see explanation under (178). The homophony of the heterographs in (185), i.e. that ق ‹q› in *A* is pronounced as *g*, is assumed on the basis that both come from works of a single author.

(183)

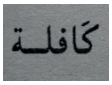
‹**q**al› (*ḤB* 4[4])

‹**ǵ**al˙› (*BB* 45$_{1}$)

gal ‘he said’ (cf. SA قال ‹qal› *qāla* ‘idem’)

(184)

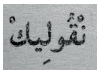
‹**q**aflẖ› (*ḤB* 20[5])



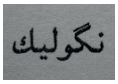
‹**ǵ**aflẖ› (*XM* 66$_{11}$)

gafla ‘caravan; a host of’ (cf. SA قافلة ‹qaflẖ› *qāfila*^{tun} ‘caravan’)

(185)



‹n˙**q**^{u}wl^{i}yk˙› (*ḌʔI* 13[5])

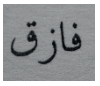


‹n**ǵ**wlyk› (*Ḍa.* 46$_{7}$)

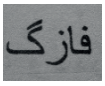
ngul-lik ‘I tell you’ (cf. SA نقول لك ‹nqwl_lk› *naqūl*^{u} *laka* ‘we tell you’)

The graphy گ ‹ǵ› is based on exogenous adaptation (see under “ج ‹ǧ› Rv گ ‹ǵ›”). For words having SA cognates with *q*, as in (183)-(185), the use of ق ‹q› can be explained either as a result of donor form-orientation or historically motivated endogenous adaptation (see “ق ‹q› Rv ڤ ‹v›”). In graphic words with no SA cognates, such as (186) and (187), ق ‹q› is an instance of analogical graphy (see, again, “ق ‹q› Rv ڤ ‹v›”).

(186)



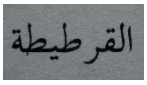
‹faz**q**› (*Mn.* 29$_{12}$)



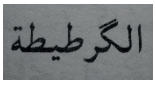
‹faz**ǵ**› (*Ṛḥ.* 142[6])

fazəg ‘wet’

(187)



‹al**q**rtyṭẖ› (*FʕI* 6[8])



‹al**ǵ**rtyṭẖ› (*MX* 171[8])

l-gərṭiṭa ‘little beast, prankster^{def}’

ق ‹q› ~ ڤ ‹ĝ›

All possible positions / Restricted variation (ق ‹q› Rv ڤ ‹ĝ›)
This variation is analogous to "ق ‹q› Rv ك ‹k›" and "ق ‹q› Rv گ ‹ǵ›", which were discussed above. The homophony of the heterographs in the first two examples, i.e. that ق ‹q› in their Forms *A* is pronounced as *g*, is assumed on the following basis: The heterographs in (188) are used in a single work, those in (190) – in works of a single author. As for (189), there is admittedly no cogent evidence that ق ‹q› is pronounced as *g* in Form *A*, but nothing rules this out.

(188)

‹**q**aflħ› (*ḤB* 20^{5})

‹**ĝ**aflħ› (*ḤB* 13_{4})

gafla 'caravan; a host of' (cf. SA قافلة ‹qaflħ› *qāfila*^{tun} 'caravan')

(189)

‹y**q**wl› (*MX* 164^{9})

‹y**ĝ**wl› (*ḌK* 9^{5})

igul 'he says' (cf. SA يقول ‹yqwl› *yaqūl*^{u} 'idem')

(190)

‹fw**q**› (*Mu.* 29^{8})

‹fw**ĝ**› (*ḤB* 16^{8})

fug 'over, upon' (cf. SA فوق ‹fwq› *fawqa* 'idem')

The graphy ڤ ‹ĝ› is based on exogenous adaptation (discussed under "ج ‹ǧ› Rv ڤ ‹ĝ›"). Since the words in (188)-(190) have SA cognates with *q*, their spelling with ق ‹q› can be interpreted either as a result of donor form-orientation or historically motivated endogenous adaptation (see "ق ‹q› Rv ڤ ‹v›"). In the following two graphic words, which have no SA cognates, the graphy ق ‹q› is analogical (see, again, "ق ‹q› Rv ڤ ‹v›"):

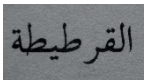
(191)

‹**q**aʕ› (*Ṣa.* 233_{10})

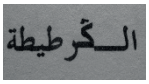
‹**ĝ**aʕ› (*ḤB* 4_{9})

gaʕ 'even', intensifying particle

(192)



‹al**q**rtyṭħ› (*Fʕ* 6^{8})



‹al**ĝ**rtyṭħ› (*ḤB* 38^{7})

l-gərṭiṭa 'little beast, prankster^{def}'

GRAPH ڤ ‹v›

This graph is used in MA graphy to mark two sounds: *v* and *g*. Not being part of the SA alphabet, it was created by adding one dot to the two-dotted graph ق ‹q›. For reasons expounded in Sect. III.1., its unconnected (non-neutralized) shape is expected to be ڢ ‹v_f› when marking *v* and ڧ ‹v_q› when marking *g*. This however, has not been confirmed by the corpus: No occurrence for ڢ ‹v_f› marking a final *v* has been recorded as this sound does not occur word-finally in MA. However, in the Arabic transcription of the French word *grave* [gʀav] ‘serious’, ending in [v], the shape is, as expected, the shallow-bowled fricative *rasm*: غراڢ ‹ġrav$_f$› (*TM* 45_{10}). Contrary to expectations, in *FB*, the same *rasm*, is used consistently to mark the word-final *g*, a plosive; see Form *B* in (177) under “ق ‹q› Rv ڤ ‹v›”. This inconsistent usage is reflected in scholarly descriptions: Aguadé (2005: 246) gives ڧ ‹v_q› as marking *v* in *Tqə.*, but according to Aguadé (2006: 255), the shape for MA in general is ڢ ‹v_f›.

ڤ ‹v› ~ ف ‹f›

Word-initially, word-medially / Restricted variation (ڤ ‹v› Rv ف ‹f›)
See ف ‹f› Rv ڤ ‹v›

ڤ ‹v› ~ ق ‹q›

All possible positions / Restricted variation (ڤ ‹v› Rv ق ‹q›)
See ق ‹q› Rv ڤ ‹v›

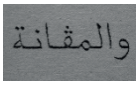
ڤ ‹v› ~ ك ‹k›

All possible positions / Restricted variation (ڤ ‹v› Rv ك ‹k›)
This variation manifests itself in graphic words in which these graphs mark the sound *g* provided it is not a reflex of SA *ǧ*. The graph ڤ ‹v› is used as a result of exogenous adaptation, while the principle underlying the use of ك ‹k› is that of phonetically motivated endogenous adaptation (see under “ج ‹ǧ› ~ ك ‹k›”). Heterograph pairs can be found for words with SA cognates, shown in (193), and without them, exemplified in (194). This variation is expected to occur in the word-final position as well.

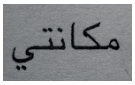
(193) 

‹**v**al› (*FB* 71[2])
gal ‘he said’ (cf. SA قال ‹qal› *qāla* ‘idem’)

‹**k**al› (*ʕD* 202$_{10}$)

(194) 

‹[wal]m**v**an[ħ]› (*FB* 70[3])
[*u-l-*]*magan*[*a*]
‘and the watch’



‹m**k**an[ty]› (*ṮN* 8[7])
magan[*ti*]
‘my watch’

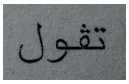
ڤ ‹v› ~ گ ‹ǵ›

All possible positions / Restricted variation (ڤ ‹v› Rv گ ‹ǵ›)
This variation manifests itself in graphic words in which these graphs mark *g* provided it is not a reflex of SA *ǧ*. Both are used in this function as a result of exogenous adaptation. Heterograph pairs can be found for words with SA cognates, shown in (196), and without them, exemplified in (195) and (197).

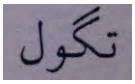
(195) 

‹**v**aʕ› (*FB* 78$_{6}$)
gaʕ ‘all’

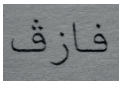
‹**ǵ**aʕ› (*ṮM* 45[10])

(196) 

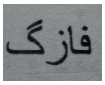
‹t**v**wl› (*FB* 70$_{9}$)
dgul ‘you say, you tell’ (cf. SA تقول ‹tqwl› *taqūl*^{u} ‘idem’)



‹t**ǵ**wl› (*ṮM* 68[8])

(197) 

‹faz**v**$_{r}$› (*FB* 6$_{2}$)
fazəg ‘wet’



‹faz**ǵ**› (*Ṛḥ.* 142[6])

ڤ ‹v› ~ ڴ ‹ĝ›

All possible positions / Restricted variation (ڤ ‹v› Rv ڴ ‹ĝ›)
This variation manifests itself in graphic words in which these graphs mark *g*, provided it is not a reflex of SA *ǧ*. It is analogous to the variation between ڤ ‹v› and گ

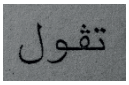
‹ǵ› discussed above. Both graphs are used in this function as a result of exogenous adaptation. Heterograph pairs can be found for words with SA cognates, shown in (199) and (200), and without them, illustrated in (198).

(198)

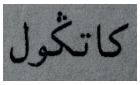
‹**v**aʕ› (*FB* 65[11])
gaʕ ‘even’, intensifying particle

‹**ĝ**aʕ› (*ḤB* 4_9)

(199)

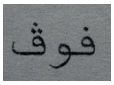


‹t**v**wl› (*FB* 70_9)
dgul ‘you say’



‹[ka]t**ĝ**wl› (*ḌK* 25_3)
[*ka-*]*dgul* ‘you are saying’
(cf. SA تقول ‹tqwl› *taqūl*^{u} ‘you say, you are saying’)

(200)



‹fw**v**$_{\mathrm{f}}$› (*FB* 77_4)
fug ‘over, upon’ (cf. SA فوق ‹fwq› *fawqa* ‘idem’)

‹fw**ĝ**› (*ḤB* 16^8)

GRAPH ك ‹k›

ك ‹k› ~ ج ‹ǧ›

Word-initially, word-medially / Restricted variation (ك ‹k› Rv ج ‹ǧ›)
See ج ‹ǧ› Rv ك ‹k›

ك ‹k› ~ غ ‹ġ›

Word-initially, word-medially / Restricted variation (ك ‹k› Rv غ ‹ġ›)
See غ ‹ġ› Rv ك ‹k›

ك ‹k› ~ ق ‹q›

All possible positions / Restricted variation (ك ‹k› Rv ق ‹q›)
See ق ‹q› Rv ك ‹k›

ك ‹k› ~ ڤ ‹v›

All possible positions / Restricted variation (ك ‹k› Rv ڤ ‹v›)

See ڤ ‹v› Rv ك ‹k›

ك ‹k› ~ گ ‹ǵ›

All possible positions: Asymmetrical variation (ك ‹k› Av گ ‹ǵ›)

This variation manifests itself in graphic words in which both graphs mark *g*, irrespective of its origin. In (201), it is a reflex of SA *q*, in (202) of *ǧ*, while (203) has no SA cognate. The graph ك ‹k› (the strong variant) is used in this function by virtue of phonetically motivated endogenous adaptation (see under "ج ‹ǧ› ~ ك ‹k›"), while the use of گ ‹ǵ› (the weak variant) is based on exogenous adaptation.

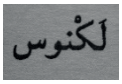
(201)

‹**k**al› (*ʕD* 202$_{10}$)

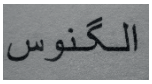
‹**ǵ**al˙› (*BB* 45$_{1}$)

gal 'he said' (cf. SA قال ‹qal› *qāla* 'idem')

(202)



‹l^{a}**k**˙nws› (*Ḍa.* 99^{6})

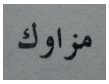


‹(a)l**ǵ**nws› (*Ṛḥ.* 154^{9})

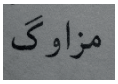
lə-gnus 'peoples, nations^{def}'

(Cf. SA الجنوس ‹alǧnws› *ʔal-ǧunūs^{u}* 'kinds^{def}')

(203)



‹mzaw**k**› (*Ḥ*B 56^{1})



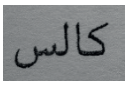
‹mzaw**ǵ**› (*MX* 135^{10})

mzawəg 'imploring'

ك ‹k› ~ ڴ ‹ĝ›

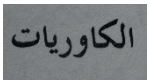
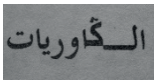
All possible positions: Asymmetrical variation (ك ‹k› Av ڴ ‹ĝ›)

This variation mirrors the preceding one: it manifests itself in graphic words in which both graphs mark *g*, irrespective of its origin. In (204), it is a reflex of SA *ǧ*, in (206) of *q*, while (205) has no SA cognate. The graph ك ‹k› (the strong variant) is used in this function by virtue of phonetically motivated endogenous adaptation (see under "ج ‹ǧ› ~ ك ‹k›"), while the use of گ ‹ǵ› (the weak variant) is based on exogenous adaptation.

(204)  

‹**k**als› (*ʕD* 111[5]) ‹**ĝ**als› (*ḤB* 21[1])

galəs ‘sitting’ (participle) (cf. SA جالس ‹ğals› *ğālis*[un] ‘idem’)

(205)  

‹al**k**awryat› (*ḤB* 42$_9$) ‹al**ĝ**awryat› (*ḤB* 14$_2$)

l-gawṛ̆yyat ‘foreign women[def]’ (despective)

(206)  

‹zr**k**› (*ḤB* 16$_3$) ‹z[w]r**ĝ**› (*ḤB* 12[8])

ẓṛəg ‘blue’ *ẓ*[*u*]*ṛəg* ‘blue[pl]’

(cf. SA أزرق ‹åzrq› *ʔazraq*[u] ‘blue’, زرق ‹zrq› *zurq*[un] ‘blue[pl]’)

GRAPH گ ‹ǵ›

گ ‹ǵ› ~ ج ‹ǧ›

Word-initially, word-medially / Restricted variation (گ ‹ǵ› Rv ج ‹ǧ›)

See ج ‹ǧ› Rv گ ‹ǵ›

گ ‹ǵ› ~ غ ‹ġ›

Word-initially, word-medially/ Restricted variation (گ ‹ǵ› Rv غ ‹ġ›)

See غ ‹ġ› Rv گ ‹ǵ›

گ ‹ǵ› ~ ق ‹q›

All possible positions / Restricted variation (گ ‹ǵ› Rv ق ‹q›)

See ق ‹q› Rv گ ‹ǵ›

گ ‹ǵ› ~ ڤ ‹v›

All possible positions / Restricted variation (گ ‹ǵ› Rv ڤ ‹v›)

See ڤ ‹v› Rv گ ‹ǵ›

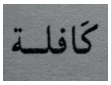
گ ‹ǵ› ~ ك ‹k›

All possible positions / Asymmetrical variation (inverse) (گ ‹ǵ› Av^{-1} ك ‹k›)
See ك ‹k› Av گ ‹ǵ›

گ ‹ǵ› ~ ڴ ‹ĝ›

All possible positions / Free variation (گ ‹ǵ› Fv ڴ ‹ĝ›)
This variation is one of the two instances of free variation which could be established on the basis of the findings in the corpus. In contrast to the other one (ى ‹á› Fv ي ‹y›), which is marginal, it is pervasive. Both graphs mark *g*, irrespective of its origin: in (207), it is a reflex of SA *q*, in (208), a reflex of SA *ǧ*, while (209) has no SA cognate. Both are used by virtue of exogenous adaptation.

(207)



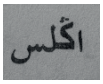
‹ǵaflẖ› (*XM* 66$_{11}$)

‹ĝaflẖ› (*Ḥ B* 13$_4$)

gafla ‘caravan; a host of’ (cf. SA قافلة ‹qaflẖ› *qāfila*^{tun} ‘caravan’)

(208)

‹aǵls› (*MX* 22^{12})



‹aĝls› (*Ḥ B* 10^4)

gləs ‘sit down!’ (cf. SA اجلس ‹ağls› *ʔiǧlis* ‘idem’)

(209)

‹drǵ› (*T̲N* 61^{10})

‹drĝ› (*Ḥ B* 10^3)

dərrəg ‘hide!’ *dərrəg* ‘he hid’[136]

(The *šadda* ّ ‹› in *A* was probably intended to be placed after ر ‹r›.)

GRAPH ڴ ‹ĝ›

ڴ ‹ĝ› ~ ج ‹ǧ›

Word-initially, word-medially / Restricted variation (ڴ ‹ĝ› Rv ج ‹ǧ›)
See ج ‹ǧ› Rv ڴ ‹ĝ›

[136] The fact that Form *A* is an imperative and Form *B* a past form does not affect pronunciation and can be disregarded here.

ڴ ‹ĝ› ~ غ ‹ġ›

{Word-initially, word-medially / Restricted variationf (ڴ ‹ĝ› Rv غ ‹ġ›)}
See غ ‹ġ› Rv ڴ ‹ĝ›

ڴ ‹ĝ› ~ ق ‹q›

All possible positions / Restricted variation (ڴ ‹ĝ› Rv ق ‹q›)
See ق ‹q› Rv ڴ ‹ĝ›

ڴ ‹ĝ› ~ ڤ ‹v›

All possible positions / Restricted variation (ڴ ‹ĝ› Rv ڤ ‹v›)
See ڤ ‹v› Rv ڴ ‹ĝ›

ڴ ‹ĝ› ~ ك ‹k›

All possible positions / Asymmetrical variation (inverse) (ڴ ‹ĝ› Av^{-1} ك ‹k›)
See ك ‹k› Av ڴ ‹ĝ›

ڴ ‹ĝ› ~ گ ‹ǵ›

All possible positions / Free variation (ڴ ‹ĝ› Fv گ ‹ǵ›)
See گ ‹ǵ› Fv ڴ ‹ĝ›

GRAPH ل ‹l›

Pseudo-variation

ل ‹l› Pv ا ‹a›

Word-initially: *See* ا ‹a› Pv ل ‹l›

Word-medially: *See* ا ‹a› Pv ل ‹l›

ل ‹l› Pv أ ‹å›

Word-initially: *See* أ ‹å› Pv ل ‹l›

Word-medially: *See* أ ‹å› Pv ل ‹l›

GRAPH م ‹m›

م ‹m› ~ ن ‹n›

Word-medially / Restricted variation (م ‹m› Rv ن ‹n›)
This variation manifests itself in graphic words in which these graphs mark *m* followed by *b*. It is marginal since only one heterograph pair could be composed:

(210)  ‹ka**m**bw› (*Ṣa.* 237[13])
kambu ‘nitwit’

 ‹ka**n**bw› (*Ḥ B* 24[1])

The graphy م ‹m› is adoptive phonetic, based on general donor graphy, while ن ‹n› can be accounted for in terms of morphological graphy: The sound *m* is interpreted as resulting from the assimilation of *n* to *b* (bilabialization). In fact, that *n* is part of the root is visible in the plural of this word: *kwanəb* (not *kwaməb**).

Graphies with ن ‹n› have also been recorded for borrowings which cannot contain a morphological *n* because their source words are pronounced with *m*. For instance, *ḅumḅa* ‘pump’ and *ṯ-ṯṛ̆mba* ‘water pump[def]’ are spelt with ن ‹n›: بونبة ‹bwnbħ› (*MX* 41$_7$) and الطرونبة ‹alṯrwnbħ› (*FB* 70$_5$), although their source words, Spanish *bomba* and Italian *tromba*, are pronounced with *m*. In such cases, the graphy ن ‹n› is probably analogical: The speller is aware that *m* followed by *b* is marked as ن ‹n› in some MA words (morphological principle) and uses this graphy in words in which this graphy is not justified by morphology. The graph ن ‹n› can also be chosen under the influence of the graphy of SA words which are written in this manner but pronounced with *m*, such as *dužambir* ‘December’ (Moroccan SA), written دجنبر ‹dğnbr› (cf. Wehr 1985: IX).

GRAPH ن ‹n›

ن ‹n› ~ ا ‹a›

Word-finally / Restricted variation (ن ‹n› Rv ا ‹a›)
See ا ‹a› Rv ن ‹n›

ن ‹n› ~ م ‹m›

Word-medially / Restricted variation (ن ‹n› Rv م ‹m›)
See م ‹m› Rv ن ‹n›

Pseudo-variation

ن ‹n› Pv ا ‹a›

Word-initially: *See* ا ‹a› Pv ن ‹n›

ن ‹n› Pv أ ‹å›

Word-initially: *See* أ ‹å› Pv ن ‹n›

GRAPH ه ‹h›

ه ‹h› ~ ا ‹a›

Word-finally / Restricted variation (ه ‹h› Rv ا ‹a›)
See ا ‹a› Rv ه ‹h›

ه ‹h› ~ ة ‹ħ›

All possible positions (word-finally) / Restricted variation (ه ‹h› Rv ة ‹ħ›)
This variation manifests itself in graphic words in which these graphs mark the word-final *a* (see "ا ‹a› Rv ة ‹ħ›" for discussion)[137].

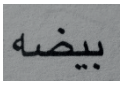
(211)

‹lyl**h**› (*Mu.* 42$_3$)

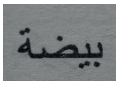
‹lyl**ħ**› (*MḤ* 37^5)

lila 'night' (cf. SA ليلة ‹lylħ› and ليله ‹lylh› *layla*^{tun} 'idem')

(212)



‹byḍ**h**› (*ṮN* 56^9)

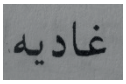
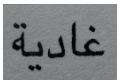


‹byḍ**ħ**› (*Ṛḥ.* 30^{10})

biḍa 'white^f' (cf. SA بيضاء ‹byḍao› *bayḍāʔ*^{u} 'idem')

---

[137] A phenomenon analogous to ي ‹y› marking *a* instead of ى ‹á›, its dotless variant, explainable in terms of analogical graphy (see "ى ‹á› Rv ي ‹y›") has not been recorded for ه ‹h› and ة ‹ħ› in the corpus, i.e. ة ‹ħ› is never used to mark *u* instead of ه ‹h›, its dotless variant.

(213)  

‹ġadyh› (*MX* 151[1]) ‹ġadyħ› (*TN* 30$_{11}$)

ġadya ‘going to[f]’ (future marker)

(cf. SA غاديه ‹ġadyh› or غادية ‹ġadyħ› *ġādiya*[*tun*] ‘becoming[f]’)

The use of ه ‹h› and ة ‹ħ› to mark *a* is based on the SA practice (see a discussion of ة ‹ħ› in Sect. I.3., point 4.e.), with necessary modifications required by MA morphosyntax. Most MA feminine nouns end in *a* in the free state (i.e. without a genitive attribute), e.g. *lila* ‘night’, with this sound being marked in various ways: usually ة ‹ħ› and ه ‹h›, but also ا ‹a› or ى ‹á›. If such a noun is used in the construct state (i.e. is followed by a genitive attribute), *a* is replaced by *t* (or, less often, *t* is added to it), e.g. *lilt l-ʕid* ‘the night of the feast’. As the corpus reveals, this opposition between *a* and *t* is reflected on the graphic level: *t* is always marked as ة ‹ħ› (in sporadic cases, ت ‹t›, see “ت ‹t› Rv ة ‹ħ›”), never as ه ‹h›[138]. Hence, *lilt l-ʕid* is written ليلة العيد ‹lylħ_alʕyd›. This is the reason why the variation between ه ‹h› and ة ‹ħ› is restricted, not asymmetrical.

Since some words, such as (213), cannot be used in the construct state, their final ه ‹h› and ة ‹ħ› always mark *a*. This means that both graphs, having been refunctionalized in this way within SA, can mark this sound irrespective of morphological factors – albeit not in verbs or pronouns, cf. footnote 110). Consequently, these two graphies are considered adoptive phonetic, based on general donor graphy (and if there is a SA cognate spelt with ه ‹h› or ة ‹ħ›, coincident with the donor graphy) rather than relativized to the donor graphy (donor-oriented or donor-defying).

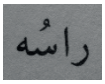
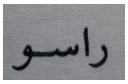
ه ‹h› ~ و ‹w›

Word-finally / Restricted variation (ه ‹h› Rv و ‹w›)

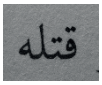
This variation manifests itself, frequently, in graphic words in which these graphs mark the pronoun *-u* ‘him, his; it[m], its[m]’ suffixed to nouns, verbs and prepositions. The graphy ه ‹h› is donor form-oriented – the cognate SA suffix *-hu/-hi* ‘idem’ is

[138] Therefore, if a graphic word with a final ه ‹h› which in a context-free position can be interpreted either as marking the nominal feminine suffix *a* or the suffixed pronoun ‘him, his; it[m], its[m]’ *-h*, e.g. سميه ‹smyh›, interpretable either as *smĭyya* ‘name’ or *səmmi-h* ‘call him’, is followed immediately by a noun, this ه ‹h› is more likely to be interpreted by native speakers as marking the suffixed pronoun *-h*, rather than the feminine suffix *-t*. For example, سميه احمد ‹smyh_aḥmd› is more likely to be interpreted as *səmmi-h Ḥməd* ‘call him Ahmad’ rather than *smĭyyt Ḥməd* ‘the name of Ahmad’.

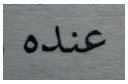
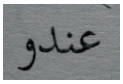
marked in this way. The graphy و ‹w›, which seems to be more frequent, especially with verbs, is adoptive phonetic, based on general donor graphy. Examples:

(214)  

‹ras^{u}**h**› (*BB* 90^{9}) ‹ras**w**› (*ḤD* 23^{8})
ṛaṣ-u ‘his head’ (cf. SA رأسه ‹råsh› *raʔsu-hu* ‘idem’)

(215)  

‹qtl**h**› (*BB* 108$_{3}$) ‹qtl**w**› (*Mu.* 38^{13})
qətl-u ‘he/it^{m} killed him’ (cf. SA قتله ‹qtlh› *qatala-hu* ‘idem’)

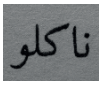
(216)  

‹ʕnd**h**› (*Mn.* 41$_{4}$) ‹ʕnd**w**› (*Mn.* 43$_{1}$)
ʕənd-u ‘he has’, lit. ‘with him, at him’ (cf. SA عنده ‹ʕndh› *ʕinda-hu* ‘idem’)

Rarely, this variation manifests itself in graphic words ending in *u* which is not this suffixed pronoun, for instance:

(217)  

‹(ã)šn**h**› (*ḌK* 30$_{9}$)[139] ‹(a)šn**w**› (*ṮN* 46^{11})
ašnu/šnu ‘what’

(218)  

‹nakl**h**› (*Mn.* 87$_{10}$) ‹nakl**w**› (*MḤ* 38$_{3}$)
naklu ‘we eat’[140]

In such cases, the reason for the graphy و ‹w› is the same as above, but ه ‹h› has to be explained as analogical: The speller is aware that ه ‹h› can mark word-final *u* in some words (donor form-orientation) and uses it in a word in which this is not justified because the word-final *u* is not the suffixed pronoun ‘him, his; it^{m}, its^{m}’[141].

[139] Recorded also by Hoogland (2013a: 71) as one of the ‘unique spellings’.

[140] In both heterographs, the reading *nakl-u* ‘I eat it^{m}’, with *u* being the suffixed pronoun – is ruled out by context.

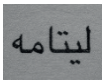
[141] An alternative, morphological and historical, interpretation of the use of ه ‹h› in (217) could be based on the fact that the interrogative *ašnu/šnu* is a result of the fusion of the interrogative *aš*

Variation between ه ‹h› and و ‹w› has not been recorded in the corpus word-medially although the morpheme *u* 'him, his; it^{m}, its^{m}' does occur in this position, in most cases, as a result of suffixing the negational *-š*. In such situations, *u* is always marked phonetically, as و ‹w›. For instance, *ma nʕəml-u-š* 'I don't do it^{m}' is spelt ما نعملوش ‹ma_nʕmlwš› (*BT* 132[7]), not *ما نعملهش ‹ma_nʕmlhš›.

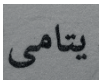
ه ‹h› ~ ى ‹á›

All possible positions (word-finally) / Restricted variation (ه ‹h› Rv ى ‹á›)
This variation manifests itself, rather rarely, in graphic words in which these graphs are used to mark the word-final *a*. Examples:

(219)



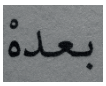
‹[l]ytam**h**› (*ṮN* 53_{10})
[*l-*]*itama* 'orphans^{def}'



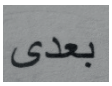
‹ytam**á**› (*ḤB* 9_{4})
itama 'orphans'

(cf. SA يتامى ‹ytamá› *yatāmā* 'idem')

(220)



‹bʕd**h̓**› (*BB* 75^{5})
bəʕda 'already; first'



‹bʕd**á**› (*Ṛḥ.* 157_{3})

In (219), for which a SA cognate exists, the graphy in Form *A* is donor form-defying, while that in Form *B* is adoptive phonetic, based on general donor graphy and, additionally, coincident with the donor graphy. In (220), which has no SA cognate, both graphies are adoptive phonetic, based on general donor graphy.

Pseudo-variation

ه ‹h› Pv ي ‹y›

Word-finally
This inter-variety pseudo-variation occurs, marginally, in graphic words which end in *i* and have SA cognates with a final *h*, marked as ه ‹h›[142]. The graphy ي ‹y›

---

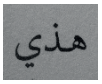
and the pronoun *hŭwa* 'he, it', with an 'n-extension' (cf. Heath 2002: 477-478, where *aš-n huwa* is adduced, among other forms). It is, however, unlikely that the speller using the graphy ه ‹h› should have this structure in mind.

[142] Cf. Heath (2002: 180) on the loss of the final *h* in MA.

marking *i* is adoptive phonetic, based on general donor graphy, while ه ‹h› is donor form-oriented. Examples:

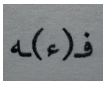
(221)

‹hḏ**h**› (*BB* 159$_8$)
hadi ‘this one^{f}’ (cf. SA هذه ‹hḏh› *hāḏihi* ‘idem’)

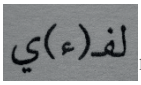


‹hḏ**y**› (*BB* 91^4)

(222)



‹f**ʔh**› (*FL* 144^8)
fʔi
‘expert in Islamic law, Quranic school teacher’
(cf. SA فقيه ‹fqyh› *faqīh^{un}* ‘expert in Islamic law’)



‹[l]f**ʔy**› (*FL* 144^5)
[*lə-*]*fʔi*
‘idem^{def}’

GRAPH ة ‹ħ›

The functions of this graph, which occurs only in the word-final position, are discussed under “ه ‹h› Rv ة ‹ħ›”; cf. also Sect. I.3, point 4.e.

ة ‹ħ› ~ ا ‹a›

All possible positions (Word-finally) / Restricted variation (ة ‹ħ› Rv ا ‹a›)
See ا ‹a› Rv ة ‹ħ›

ة ‹ħ› ~ ت ‹t›

All possible positions (Word-finally) / Restricted variation (ة ‹ħ› Rv ت ‹t›)
See ت ‹t› Rv ة ‹ħ›

ة ‹ħ› ~ ه ‹h›

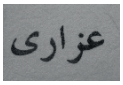
All possible positions (Word-finally) / Restricted variation (ة ‹ħ› Rv ه ‹h›)
See ه ‹h› Rv ة ‹ħ›

[143] See Sect. II.2. for explanation of the use of parentheses, irrelevant in this case.

All possible positions (Word-finally) / Restricted variation (ة ‹ħ› Rv ى ‹á›)
This variation manifests itself in graphic words representing nouns, adjectives or adverbs, with the exclusion of verbs and pronouns[144], in which these graphs mark a word-final *a*. Both graphies are adoptive phonetic, based on general donor graphy (within which ة ‹ħ› has been refunctionalized to mark this vowel). In (223) and (224), graphic words with no SA cognates are exemplified.

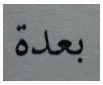
(223)

‹ʕzarħ› (*Ṣa.* 213[6])
ʕzara ‘bachelors’

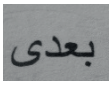


‹ʕzará› (*Ḥb* [2]63)

(224)



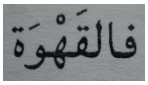
‹bʕdħ› (*Mn.* 45[5])
bəʕda ‘already; first’



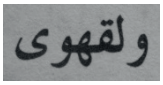
‹bʕdá› (*Ṛḥ.* 157_3)

If the word has a SA cognate spelt with ة ‹ħ›, as exemplified in (225) and (226), the graphy ى ‹á› is donor form-defying. If the SA cognate is spelt with ى ‹á›, as in (227), it is ة ‹ħ› that is based on this spelling principle.

(225)



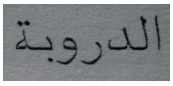
‹[fa]lqᵃhʿwᵃħ› (*Mm.* 119_9)
[*fə-*]*l-qəhwa* ‘[in] the café’



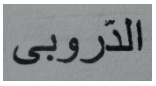
‹[w]lqhwá› (*Mu.* 37_{14})
[*u-*]*l-qəhwa* ‘[and] the café’

(cf. SA القهوة ‹alqhwħ› *ʔal-qahwa*^{tu} ‘coffee; café^{def}’)

(226)



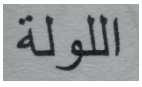
‹aldrwbħ› (*FB* 78[5])
d-druba ‘small streets^{def}’



‹aldˁrwbá› (*Ṛḥ.* 198[7])

(cf., for instance, السبوعة ‹alsbwʕħ› *ʔas-subūʕa*^{tu} ‘predatory animals^{def}’[145])

(227)



‹allwlħ› (*Ṛḥ.* 148_4)
l-lŭwwla ‘first$^{f\ def}$’ (cf. SA الأولى ‹alåwlá› *ʔal-ʔūlā* ‘idem’)

[144] For an isolated case of a verb written with the word-final ة ‹ħ›, see footnote 110.
[145] See example (60) in this chapter for explanation of referring to the morphological cognacy.

GRAPH و ‹w›

و ‹w› ~ أ ‹å›

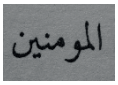
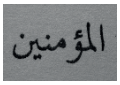
Word-initially / Restricted variation (و ‹w› Rv أ ‹å›)
See أ ‹å› Rv و ‹w›

و ‹w› ~ ه ‹h›

Word-finally / Restricted variation (و ‹w› Rv ه ‹h›)
See ه ‹h› Rv و ‹w›[146]

و ‹w› ~ ۇ ‹ẘ›

Word-medially / Restricted variation (و ‹w› Rv ۇ ‹ẘ›)
This variation manifests itself, rather rarely, in graphic words in which these two graphs mark *u* or *w* corresponding to *ʔ* in SA cognates spelt with ۇ ‹ẘ›. The variation is restricted, not asymmetrical, because و ‹w› cannot be substituted for ۇ ‹ẘ› in graphic words, usually representing borrowings from SA, in which the latter graph marks *ʔ*: For example, *suʔal* ‘question’ is consistently written سؤال ‹sẘal›, not *سوال ‹swal›. The graphy و ‹w› marking *u* or *w* is adoptive phonetic, based on general donor graphy, while ۇ ‹ẘ› is donor form-oriented. Example:

(228)  

‹alm**w**mnyn› (*MḤ* 20[13]) ‹alm**ẘ**mnyn› (*MḤ* 26[5])
l-mumnin ‘believers[the]’ (cf. SA المؤمنين ‹almẘmnyn› *al-muʔminīn*[a] ‘idem[obl]’)

Although it cannot be ruled out that the pronunciation of *B* intended by the speller was the donor-oriented pronunciation with *ʔ*, i.e. *l-muʔminin*, it seems justified to assume that in this and similar cases, the graphy with ۇ ‹ẘ› also allows the more colloquial pronunciation *l-mumnin*.

[146] A possible but not recorded restricted word-medial variation between و ‹w› and ه ‹h› is briefly discussed there.

Pseudo-variation

و ‹w› Pv أ ‹å›

Word-initially: *See* أ ‹å› Pv و ‹w›

و ‹w› Pv آ ‹ã›

Word-medially: *See* آ ‹ã› Pv و ‹w›

GRAPH ۇ ‹ẘ›

In the corpus, the graph ۇ ‹ẘ› is recorded only in the word-medial position. However, it has been described by other scholars (Hoogland 2013a: 71; Mion 2014: 193) as marking the conjunction *w-/u-* 'and' in *Ami.* (not included in the corpus), a function in which this graph occurs word-initially.

ۇ ‹ẘ› ~ أ ‹å›

Word-medially / Restricted variation (ۇ ‹ẘ› Rv أ ‹å›)
See أ ‹å› Rv ۇ ‹ẘ›

ۇ ‹ẘ› ~ و ‹w›

Word-medially / Restricted variation (ۇ ‹ẘ› Rv و ‹w›)
See و ‹w› Rv ۇ ‹ẘ›

ۇ ‹ẘ› ~ ئ ‹ÿ› and ۇ ‹ẘ› ~ ء ‹o›

These two variations can be expected in graphic words representing borrowings from SA that have retained *ʔ* in MA pronunciation. Since such graphic words are infrequent, these variations would be marginal.

{Word-medially / Restricted variation (ۇ ‹ẘ› Rv ئ ‹ÿ›)}
For graphic words such as مسؤول ‹msẘwl› *məsʔul* 'responsible' (*Mu.* 7_6) orميئوسة ‹myÿwsħ› *məyʔusa* (*Ḍḥ.* 163[10]) 'desperate[f]', heterographs with ئ ‹ÿ›, viz. مسئول ‹msÿwl› and ميؤوسة ‹myẘwsħ›, can be expected on the basis of the occurrence of such forms in written Moroccan SA.

{Word-medially / Restricted variation (ؤ ‹ẘ› Rv ء ‹o›)}
For graphic words such as الرؤوف ‹alrẘwf› *r-raʔuf* ‘compassionate^{def}’ (part of a proper name) (*XM* 148[2]) or دؤوب ‹dẘwb› *daʔub* ‘tireless, persevering’ (*Ṛḥ.* 119_8), heterographs with ء ‹o›, viz. الرءوف ‹alrowf› and دءوب ‹dowb›, can be expected on the basis of the occurrence of such forms in written Moroccan SA.

GRAPH ى ‹á›

The graph ى ‹á› occurs only word-finally.

ى ‹á› ~ ا ‹a›

All possible positions (word-finally) / Restricted variation (ى ‹á› Rv ا ‹a›)
See ا ‹a› Rv ى ‹á›

ى ‹á› ~ أ ‹å›

{All possible positions (word-finally) / Restricted variation (ى ‹á› Rv أ ‹å›)
See أ ‹å› Rv ى ‹á›

ى ‹á› ~ ه ‹h›

All possible positions (word-finally) / Restricted variation (ى ‹á› Rv ه ‹h›)
See ه ‹h› Rv ى ‹á›

ى ‹á› ~ ة ‹ħ›

All possible positions (word-finally) / Restricted variation (ى ‹á› Rv ة ‹ħ›)
See ة ‹ħ› Rv ى ‹á›

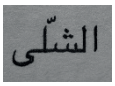
ى ‹á› ~ ي ‹y›

All possible positions (word-finally) / Free variation (ى ‹á› Fv ي ‹y›)
This very specific variation manifests itself, marginally, in two situations: (i) when these graphs are used to mark a word-final *a* or *i*/*y* and (ii) regularly but in only one source and one abstract graphic word, when they mark no sound.

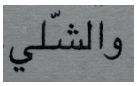
As for the first situation, the use of ى ‹á› and ي ‹y› to mark *a* or *i/y* is adoptive phonetic, based on general donor graphy. However, while marking *i/y* is the typical function of ي ‹y› in general SA, Moroccan SA and MA, this is not the case

with ى ‹á›, which is used in this way only by some spellers. For instance, SA *fī* 'in', apart from having the graphic form في ‹fy›, can also be written فى ‹fá›. This practice, known "in many regions" (Badawi *et al.* 2004: 15, 20), is especially widely used in Egypt – for both Egyptian SA and EA – where ى ‹á› used in this way is referred to as *al-yāʔ al-miṣriyya* 'Egyptian *yāʔ*' (as opposed to *al-yāʔ al-šāmiyya* 'Levantine *yāʔ*, i.e. ي ‹y› marking *i*) (cf. also footnote 23 on its name). The corpus reveals that the extent of this practice in MA is marginal: Only two such graphies have been recorded. They are shown as Forms *A* in (229) and (230).

(229)

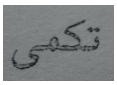


‹alš·l**á**› (*ŢN* 45[5])
š-šli 'chairs[def]'

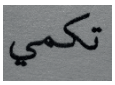


‹[w]alš·l**y**› (*ŢN* 45[5])
[*u*-]*š-šli* '[and] chairs[def]'

(230)



‹tkm**á**› (*BT* 68[1])
təkmi 'you smoke'

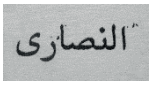


‹tkm**y**› (*MḤ* 17[5])

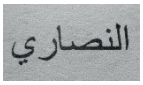
Apart from these two graphic words, *i* is consistently marked as ي ‹y› and perhaps the two cases should be treated as misprints. However, since this graphy can also be found, albeit sporadically, in Moroccan SA texts, it is possible to claim that these forms are a sign of a spelling tendency which has not penetrated into the corpus substantially. In other words, although on the abstract level ى ‹á› can be used instead of the word-final ي ‹y› in any graphic word, it is only exceptionally put into practice on the level of actual graphic words.

If the use of ى ‹á› instead of ي ‹y› to mark *i/y* in MA is not quite clear, the use of ي ‹y› instead ى ‹á› to mark *a* is even more so: Consider the following pairs that have been chosen out of a dozen established on the basis of occurrences recorded in the corpus:

(231)

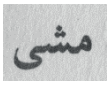


‹alnṣar**á**› (*QQ* 49[4])
n-nṣaṛa 'Christians, Westerners[def]' (cf. SA النصارى ‹alnṣará› *ʔan-naṣārā* 'idem')

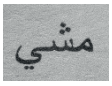


‹alnṣar**y**› (*Ṛḥ.* 169_7)

(232)



‹mš**á**› (*Mu.* 14[3])
mša 'he/it[m] went' (cf. SA مشى ‹mšá› *mašā* 'idem')



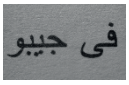
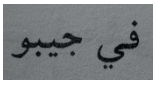
‹mš**y**› (*Ṛḥ.* 168_9)

(233)  

‹yl**á**› (*ḌK* 8_2) ‹yl**y**› (*ṮN* 53^3)
ila ‘if’

The graphy ى ‹á› is adoptive phonetic, based on general donor graphy, in (231) and (232), additionally, coincident with the donor graphy, but the reason behind the use of ي ‹y› is not obvious. It mirrors a phenomenon observed in SA, where ي ‹y› at times marks the long final *ā* – a spelling which, strictly speaking, is an orthographic error[147]. The reason behind it seems to be analogical graphy: The speller is aware that ى ‹á› and ي ‹y› are mutually substitutable in some words word-finally and uses ي ‹y› in a word in which there is no sound that would justify this. However, this analogy-based explanation is not very convincing in the case of MA graphy, in which the basis for the analogy is weak: What should induce spellers to use ي ‹y› instead of ى ‹á› for marking *a*, if ى ‹á› is not very often used instead of ي ‹y› for marking *i*, remains an open question.

The second situation distinguished at the beginning of the section devoted to the present variation is a particular case too. It manifests itself in graphic words representing the preposition *f-* ‘in’, which means that these graphs mark no sound. Example:

(234)  

‹f**á**_ğybw› (*Ṛḥ.* 72^6) ‹f**y**_ğybw › (*MX* 130_9)
f-žib-u ‘in his pocket’
(cf. SA في جيبه ‹fy_ğybh› and فى جيبه ‹fá_ğybh › *fī ğaybi-hi* ‘idem’)

The graphy ى ‹á› in this function has been recorded in one source only, *Ṛḥ.*, but is used quite consistently there. The graphy ي ‹y›, by contrast, is frequent in other sources[148]. Both graphies are clearly donor form-oriented.

---

[147] Badawi *et al*. (2004: 20) regard it as one of the “many typographical inconsistencies” in modern SA graphy which indicate that the system is still unstable. Buckwalter (2004: 32) observes this phenomenon in Arabic newswire texts and says that “[i]t is not entirely clear whether these “dotted” *alif maqsura*’s were produced by human typists or by an encoding conversion process gone awry”. He proposes the following explanation: “It is possible that the original keyboarding was done on a platform where word-final *ya’* [ي ‹y› – M.M.] and *alif maqsura* [ى ‹á› – M.M.] are displayed via visually identical “un-dotted” glyphs, so it makes no difference which of the two keys the typist presses on the keyboard: both produce the same visual display, but are stored electronically as two different characters”. It should be added that on some electronical devices, e.g. some mobile phones, the graph ى ‹á› is entered indirectly, by selecting it from a list opened through the key ي ‹y›, which makes it more time-consuming to type than ي ‹y›. This obviously does not explain why this phenomenon occurs in printed literary texts and news texts.

[148] Apart from these two graphies, this preposition is usually written as ف ‹f_i› followed by a space or ڤ ‹f› connected to a following word.

GRAPH ي ‹y›

ي ‹y› ~ ا ‹a›

Word-initially / Restricted variation (ي ‹y› Rv ا ‹a›)
See ا ‹a› Rv ي ‹y›

Word-medially / Restricted variation (ي ‹y› Rv ا ‹a›)
See ا ‹a› Rv ي ‹y›

Word-finally / Restricted variation (ي ‹y› Rv ا ‹a›)
See ا ‹a› Rv ي ‹y›

ي ‹y› ~ إ ‹ạ›

Word-initially / Restricted variation (ي ‹y› Rv إ ‹ạ›)
See إ ‹ạ› Rv ي ‹y›

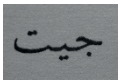
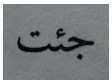
Word-medially / Restricted variation (ي ‹y› Rv إ ‹ạ›)
See إ ‹ạ› Rv ي ‹y›

ي ‹y› ~ ى ‹á›

All possible positions (word-finally) / Free variation (ي ‹y› Fv ى ‹á›)
See ى ‹á› Fv ي ‹y›

ي ‹y› ~ ئ ‹ỷ›

All possible positions (word-medially, word-finally) / Restricted variation (ي ‹y› Rv ئ ‹ỷ›)
This variation manifests itself in graphic words representing words that have SA cognates with *ʔ* spelt as ئ ‹ỷ› to which *i* or *y* corresponds in MA. The graphy ي ‹y› is adoptive phonetic, based on general donor graphy, while ئ ‹ỷ› is donor form-oriented. Examples:

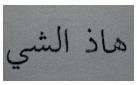
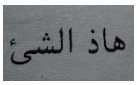
(235)  

‹ğyt› (*ṮM* 9[11]) ‹ğẙt› (*BB* 43[7])
žit ‘I came’ (cf. SA جئت ‹ğẙt› *ğiʔtu* ‘idem’)

(236)  

‹dayma› (*Bə.* 71$_2$) ‹daẙma› (*Ḥ.* 79$_9$)[149]
daymən ‘always’ (cf. SA دائما ‹daẙma› *dāʔiman* ‘idem’)

Word-finally, this variation is marginal as it is manifested in one heterograph pair only (note that ئ ‹ẙ› does not mirror the SA graphy faithfully):

(237)  

‹had_alšy› (*Aʕ* 33$_2$) ‹had_alšẙ› (*ḌK* 22$_3$)
had š-ši ‘this’, lit. ‘this thing[def]’
(cf. SA هذا الشيء ‹hḏa_alšyo› *hāḏā š-šayʔᵘ* ‘this thing[def]’)

The variation is restricted for the following reasons: Some graphic words with ئ ‹ẙ› are likely to have been intended by the spellers to have a pronunciation with *ʔ*, i.e. that of borrowings from SA with a meaning different from their MA cognates. For instance, the graphic words القارئ ‹alqarẙ› (*Ḥ.* 94$_4$) and كاتفاجئ ‹katfağẙ› (*Ḥ.* 99[2]) seem to have been intended to be pronounced *l-qaṛəʔ* and *ka-tfažəʔ* and mean ‘reader[def]’ and ‘you surprise’, respectively, rather than *qaṛi* and *ka-tfaži*, which mean ‘educated, having learned; having read’ and ‘you console’. Such examples suggesting that the pair ي ‹y› and ئ ‹ẙ› can be meaning-differentiating are, however, scanty. Finally, there are examples in which only *ʔ* can be pronounced and only ئ ‹ẙ› can be written, for instance, *l-yiʔs* ‘despair’, spelt اليئس ‹alyẙs› (*Ḥ.* 201[11]), cannot be spelt *اليييس ‹alyys›.

Pseudo-variation

ي ‹y› Pv ه ‹h›

Word-finally: *See* ه ‹h› Pv ي ‹y›

[149] Although Form *B*, written with ئ ‹ẙ›, can be pronounced in the SA-oriented way *daʔimən*, i.e. with *ʔ*, this does not rule out the pronunciation *daymən*.

GRAPH ئ ‹ẙ›

This graph occurs only word-medially and word-finally.

ئ ‹ẙ› ~ أ ‹å›

Word-medially / Restricted variation (ئ ‹ẙ› Rv أ ‹å›)
See أ ‹å› Rv ئ ‹ẙ›

{Word-finally / Restricted variation (ئ ‹ẙ› Rv أ ‹å›)}
See أ ‹å› Rv ئ ‹ẙ›

ئ ‹ẙ› ~ ۇ ‹ẘ›

{Word-medially / Restricted variation (ئ ‹ẙ› Rv ۇ ‹ẘ›)}
See ۇ ‹ẘ› Rv ئ ‹ẙ›

ئ ‹ẙ› ~ ي ‹y›

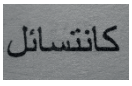
Word-medially / Restricted variation (ئ ‹ẙ› Rv ي ‹y›)
See ي ‹y› Rv ئ ‹ẙ›

Word-finally / Restricted variation (ئ ‹ẙ› Rv ي ‹y›)
See ي ‹y› Rv ئ ‹ẙ›

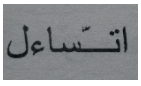
ئ ‹ẙ› ~ ۆ ‹o›

Word-medially / Restricted variation (ئ ‹ẙ› Rv ۆ ‹o›)
This variation manifests itself, marginally, in graphic words representing borrowings from SA in which these graphs mark *ʔ*. Examples:

(238)



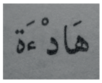
‹[kan]tsaẙl› (*Ṛḥ.* 148[1])
[*ka-nə*]*tsaʔəl* ‘I wonder’



‹[a]tˑsa**o**l› (*Ṛḥ.* 187[3])
[*tə*]*tsaʔəl* ‘you wonder’

(cf. SA نتساءل ‹ntsaol› *natasāʔal*^{u} ‘we wonder’ and تتساءل ‹ttsaol› *tatasāʔal*^{u} ‘you wonder’)

(239) 



‹hadẙħ› (*Ṛḥ.* 132^{8}) ‹h^{a}ad˙**o**^{a}ħ› (*MX* 137^{12})

hadʔa ‘calm^{f}’ (cf. SA هادئة ‹hadẙħ› *hādiʔa*^{tun} ‘idem’)

In (238), the graphy ء ‹o› is donor form-oriented. The graphy ئ ‹ẙ› seems to be donor form-defying at first glance but can also be interpreted as donor principle-oriented: If so, it is a result of applying to MA phonetic content the SA orthographic principle according to which *ʔ* preceded by *ā* and followed by *i* is marked as ئ ‹ẙ› (with MA *ə* being treated as *i*). In (239), the situation is different: ئ ‹ẙ› is donor form-oriented, while ء ‹o› is donor form-defying (as it cannot be justified by the donor-graphy or any SA orthographic principle).

GRAPH ء ‹o›

This graph occurs only word-medially and word-finally.

ء ‹o› ~ أ ‹å›

{Word-finally / restricted variation (ء ‹o› Rv أ ‹å›}

See أ ‹å› Rv ء ‹o›

ء ‹o› ~ ؤ ‹ẘ›

{Word-medially / Restricted variation (ء ‹o› Rv ؤ ‹ẘ›)}

See ؤ ‹ẘ› Rv ء ‹o›

ء ‹o› ~ ئ ‹ẙ›

Word-medially / Restricted variation (ء ‹o› Rv ئ ‹ẙ›)

See ئ ‹ẙ› Rv ء ‹o›

VI.2. Discussion

By identifying heterograph pairs and the variants of particular graphs as well as spelling principles underlying particular graphies in the previous section, questions (1) and (2) formulated in the introduction to this book, concerning elements showing variation and reasons for it, were answered as far as its qualitative type is

concerned. In what follows, questions (3) and (4) will be approached: that about qualitative invariants and phonetic ambiguity. Next, the second axis of this study indicated in the introduction will be addressed, by enumerating and discussing different types of variants of particular graphs and establishing a list of MA graphemes with their allographs.

VI.2.1. Qualitative invariants

Once all graphs affected by variation and those which have no variants have been identified, it becomes possible to determine which MA graphic words are qualitative invariants, i.e. have no qualitative variational heterographs, not only as revealed by the corpus but also as can be predicted. First, graphic words composed solely of graphs which have no variants, i.e. ح ‹ḥ›, خ ‹x›, ر ‹r›, ش ‹š› and ع ‹ʕ›, can be safely considered qualitative invariants. Examples:

عش ‹ʕš› *ʕəšš* ‘nest’,
شرع ‹šrʕ› *šṛəʕ* ‘religious law, shariah’,
عشر ‹ʕšr› *ʕšəṛ* ‘ten’,
شرح ‹šrḥ› *šṛəḥ* ‘he explained’,
شخر ‹šxr› *šxəṛ* ‘he snored’,
رش ‹rš› *ṛəšš* ‘he sprinkled’,
حر ‹ḥr› *ḥŭṛṛ* ‘free’
خرخر ‹xrxr› *xəṛxəṛ* ‘he rattled’.

The list of qualitative invariants is not limited to such graphic words. It can be claimed that it also includes graphic words representing any phonetic word which fulfils all of the following conditions:
a. it contains no emphatics and has no lexical SA cognates with emphatics, except for emphatics that have no special graphs in SA: *ḍ, ṛ*, *ḷ* and *ṃ*,
b. it contains no lexical SA cognates with interdental fricatives,
c. it contains no *g*, *p*, *f*, *q* or *v*,
d. it does not end in vowels, especially *a* (except for verbs ending in *-u* marking the plural),
e. it does not begin in a vowel or semivowel nor is a result of graphic prefixation.
Thus, qualitative invariants include, for instance:

كتاب ‹ktab› *ktab* ‘book’,
راجل ‹raǧl› *ṛažəl* ‘man’,
عبد الكريم ‹ʕbd_alkrym› *ʕəbd əl-krim*, a male given name.

Admittedly, one can imagine that one comes across graphic words with analogical graphies which can appear to be possible heterographs of such graphic words, viz. *كثاب ‹kṯab› for *ktab* or *عبذ الكريم ‹ʕbḏ_alkrym› for *ʕabd al-krim*. Such forms, however, should be considered special cases due to their irregular and marginal character.

A qualitative invariant can show other types of variational heterography: quantitative, e.g. اكتاب ‹aktab›, رجل ‹rğl›, or linear, e.g. عبدالكريم ‹ʕbdalkrym›, respectively. Words which have no qualitative, quantitative or linear heterographs are *absolute graphic invariants*. Identifying them will be possible once quantitative and linear heterography has been described in real texts.

VI.2.2. Phonetic ambiguity of graphs

In many cases, variation is related to the phonetic ambiguity of graphs. A strict view on phonetic ambiguity is taken here, in which only graphs used on the basis of the phonetic principle are taken into consideration, with the exclusion of graphies based on the morphological principle, analogical graphy and the principle of donor-orientation[150]. Graphs which the corpus has revealed to be phonetically ambiguous are listed below with the sounds they mark.

ا ‹a›	*a*, *u*, *i*, *ʔ*, *y*
أ ‹å›	*a*, *u*, *i, ʔ*
آ ‹ã›	*a*, *ʔa* (word-medially)
إ ‹ą›	*ǐ*, *i*, *y*
ب ‹b›	*b*, *ḅ, p*
ج ‹ğ›	*ž*, *g*
ر ‹r›	*r*, *ṛ*
ز ‹z›	z, *ẓ*
غ ‹ġ›	*ġ*, *g*
ف ‹f›	*f*, *q*, *v*
ق ‹q›	*q*, *g*
ڤ ‹v›	*v*, *g*
ك ‹k›	*k*, *g*
ل ‹l›	*l*, *ḷ*

[150] If the non-phonetic functions of graphs, i.e. those based on the morphological principle (e.g. ت ‹t› marking *d*, *ṭ* and *ḍ*) or the principle of donor-orientation (e.g. ت ‹t› and ث ‹ṯ› marking *ṭ*, د ‹d› marking *ḍ*, س ‹s› marking *ṣ*), were also taken in to consideration, the number of graphs characterized by ambiguity, which in this case would be functional, would be higher. The graphs ا ‹a› and أ ‹a› are special in this respect because apart from being phonetically ambiguous, they also have non-phonetic functions (signalling clusters and pseudo-clusters).

م ‹m›	*m*, *ṃ*
ه ‹h›	*h*, *a* (word-finally), *u* (word-finally)
ة ‹ẖ›	*t*, *a*
و ‹w›	*u*, *w*
ي ‹y›	*i*, *y*, *a* (word-finally)

Thus, out of a total 42 graphs, 19, i.e. nearly half, are phonetically ambiguous. However, some graphies responsible for the phonetic ambiguity are marginal. This is the case of غ ‹ġ› marking *g*, ڤ ‹f› marking *q* and ي ‹y› marking *a* word-finally.

In some cases, phonetic ambiguity occurs only in particular positions in the graphic word. For example, the phonetic function of آ ‹ã› is indeterminate when this graph is considered out of context, but within a graphic word it is determinable as marking *a* word-initially and *ʔa* word-medially. Similarly, ه ‹h› is phonetically ambiguous when used word-finally, but word-initially and word-medially it always marks *h*.

Efforts to remedy the undesirable feature of the phonetic ambiguity of some graphs have been undertaken by some spellers. They consist in using, with greater or lesser consistency, special univocal graphs introduced by way of exogenous adaptation: پ ‹p› for *p*, گ ‹ǵ›, ڴ ‹ĝ› and ڤ ‹v› for *g* as well as ڤ ‹v› for *v* (paradoxically, ڤ ‹v› itself turns out to be phonetically ambiguous). On the other hand, no attempts have been observed in the corpus to introduce special graphs that would reflect the difference between plain (non-pharyngealized) and emphatic (pharyngealized) sounds: In all sources, ب ‹b› marks both *b* and *ḅ*, ر ‹r› marks both *r* and *ṛ*, ز ‹z› marks both *z* and *ẓ*, ل ‹l› marks both *l* and *ḷ*, and م ‹m› marks both *m* and *ṃ*. The reason seems to be that if there are any minimal pairs showing that some of these pairs of sounds are bound by phonological opposition, they are so rare that marking this in writing is superfluous from the practical point of view[151].

VI.2.3. Variants and graphemes in Written Moroccan Arabic

A complete inventory of variants of a graph would be composed of its free variants, asymmetrical variants, restricted variants and pseudo-variants. This is exemplified in Table 3 for ا ‹a›, the graph with the highest number of variants, as an example (*i* = word-initially; *m* = word-medially; *f* = word-finally; ~ = marginal variation; expected variants are not indicated).

[151] Proposals for distinguishing graphically at least some of these sounds can be, however, found in scholarly literature. For instance, al-Midlāwī (2001: 168-169) proposes using Arabic letters modified by the addition of three dots for marking emphatic sounds such as *ṛ*, *ẓ* and *ẓ̌*, while Durand (2004: 52) uses a similar method (addition of three dots or a superscript *hamza*) to mark *ṛ*, *ẓ* and *ṃ*.

Table 3. *Variants of the graph* ا ‹a›

ا ‹a›	أ ‹å›	آ ‹ã›	إ ‹ą›	ت ‹t›	ل ‹l›	ن ‹n›	ه ‹h›	ة ‹ħ›	ى ‹á›	ي ‹y›
Asymmetrical variants	*i, f*	*i*	*i, m*							
Restricted variants	*m*	*m*				*f* ~	*f*	*f*	*f*	*i, m, f* ~
Pseudo-variants				*i, m*	*i, m*	*i*				

Giving such complete sets for each graph would certainly show the complexity of the MA graphic system tellingly but it seems that it would not be very illuminating regarding its structure, while taking up too much space. Let us instead take a look at a different type of data emerging from the analysis. The number of variations identified is 101 (without counting expected variations), and includes: 2 free variations, 10 asymmetrical variations and 89 restricted variations. 10 pseudo-variations have been identified: 6 of the intra-variety type and 4 of the inter-variety type. Five graphs show no variation at all.

Frequent variations should be distinguished from marginal ones. For instance, the free variation between گ ‹ģ› and ڴ ‹ĝ› is very frequent, while the other one, between ى ‹á› and ي ‹y› is marginal. All asymmetrical variations are frequent. Out of restricted variations, 75 are frequent, while 14 are marginal. Thus, apart from the special case of ى ‹á› Fv ي ‹y›, all instances of marginal variation belong to restricted variation. This means that asymmetrical variation and, to a lesser extent, free variation, are relations very well represented in the corpus, and this stability can be interpreted as proving that choosing them as a criterion for classifying two graphs into one grapheme is a right decision, in contrast to restricted variation, which may manifest itself in very few instances for some pairs of graphs.

In what follows, the variants of each variation type are listed separately for each graph. In parallel to this, particular variations are discussed in the light of the criterion proposed at the end of Sect. III.2.2.2. Finally, a list of MA graphemes with their allographs will be established.

1. **Free variation**. There is only one pair of graphs bound by free variation which is not marginal: گ ‹ģ› Fv ڴ ‹ĝ›. It manifests itself in all positions. There is no doubt that these two graphs can be classified into one grapheme. Neither of them can be considered the basic allograph. The former, however, is used in several sources included in the corpus, while the use of the latter has been recorded in two only (*ḤB* and *ḌK*).

The other pair of graphs considered to be bound by free variation is ى ‹á› and ي ‹y›. It manifests itself word-finally, i.e. in all possible positions in which ى ‹á› can occur. However, since it is represented by very few pairs of actual graphic words, it is considered marginal. Consequently, the two graphs are not classified into one grapheme.

2. **Asymmetrical variation**. In most cases, this variation manifests itself in all positions or, in one case, إ ‹ą›, in all possible positions (word-initially and word-medially). In addition, no pair bound by asymmetrical variation is marginal. All this points to its regular and structural character. Any two graphs bound by it, as listed below, are classified into one grapheme.

ا ‹a›	Av	إ ‹ą›
ب ‹b›	Av	پ ‹p›
ت ‹t›	Av	ث ‹ṯ›
د ‹d›	Av	ذ ‹ḏ›
ض ‹ḍ›	Av	ظ ‹ẓ›
ك ‹k›	Av	گ ‹ǵ›
ك ‹k›	Av	ڴ ‹ĝ›

In asymmetrical variation, the strong variant is the basic allograph and the weak one – the non-basic allograph. It is conspicuous that except for the pair ض ‹ḍ› Av ظ ‹ẓ›, the strong variants are graphically less complex than the weak ones, i.e. they are bare *rasm*s, in contrast to *rasm*s modified by the addition of a *hamza* sign, dots or strokes. This allows us to formulate the following generalization regarding the nature of a part of the domain of qualitative variation: Bare *rasm*s are able to be substituted for graphically complex forms, but not always vice versa.

Graphs bound by asymmetrical variation manifesting itself only in some positions are not necessarily classified into one grapheme. These are the following pairs:

ا ‹a›	Av	أ ‹å›, word-initially, word-finally
ا ‹a›	Av	آ ‹ã›, word-initially
أ ‹å›	Av	آ ‹ã›, word-initially

As far as these pairs are concerned, situations in which the strong variant cannot be substituted for the weak one in a given position are fairly rare. In other words, their variability is much stronger than their distinctive potential. The latter, however, cannot be ignored.

3. **No variation**. Graphs that show *no variation* are five: ح ‹ḥ›, خ ‹x›, ر ‹r›, ش ‹š›, ع ‹ʕ›. Each of them is placed into a separate grapheme, with no allographs. Their invariance is closely related to the fact that, apart from ر ‹r›, they are phonetically univocal. As for ر ‹r›, it is phonetically ambiguous, marking both plain (non-pharyngealized) *r* and emphatic (pharyngealized) *ṛ*, two sounds out of which each has the status of a phoneme (Kjamilev 1968: 24; Aguadé 2003: 78). However, since the distinctive potential of this pair manifests itself in a very limited number of words, the spellers do not feel the need to reflect it in writing.

4. **Restricted variation**, the most frequent type of variation, binds many pairs of graphs. Exemplified and described in detail in the preceding sections, its particular pairs are represented synthetically in the following tables, this time without cross-referencing, i.e. every restricted variation is indicated for every graph separately. The plus sign + means variations manifesting themselves (comparatively) frequently. The tilde ~ marks marginal variations. (Expected variations are indicated, as {}, only for variations manifesting themselves in all positions.)

ا ‹a›:

	أ ‹å›	آ ‹ã›	ن ‹n›	ه ‹h›	ة ‹ħ›	ى ‹á›	ي ‹y›
word-initially							+
word-medially	+	+					+
word-finally			~	+	+	+	~

أ ‹å›:

	ا ‹a›	آ ‹ã›	إ ‹ą›	و ‹w›	ؤ ‹ẘ›	ئ ‹ŷ›
word-initially			+	+		
word-medially	+	+			~	~
word-finally						

آ ‹ã›:

	ا ‹a›	أ ‹å›
word-medially	+	+

إ ‹ą›:

	أ ‹å›	ي ‹y›
word-initially	+	+
word-medially		+

ت ‹t›:

	د ‹d›	ط ‹ṭ›	ة ‹ħ›
word-initially	+	~	
word-medially	~	~	
word-finally		{}	+

ث ‹t̲›:

	ط ‹ṭ›
word-medially	~

ج ‹ǧ›:

	ك ‹k›	گ ‹ǵ›	ڴ ‹ĝ›
word-initially	~	+	~
word-medially	{}	+	~

د ‹d›:

	ت ‹t›	ض ‹ḍ›
word-initially	+	+
word-medially	~	+
word-finally		+

ز ‹z›:

	ص ‹ṣ›
word-finally	~

س ‹s›:

	ص ‹ṣ›
word-initially	+
word-medially	+
word-finally	{}

ص ‹ṣ›:

	ز ‹z›	س ‹s›
word-initially		+
word-medially		+
word-finally	~	{}

ض ‹ḍ›:

	د ‹d›	ط ‹ṭ›
word-initially	+	
word-medially	+	
word-finally	+	~

ط ‹ṭ›:

	ت ‹t›	ث ‹t̲›	ض ‹ḍ›	ظ ‹ẓ›
word-initially	~			
word-medially	~	~		
word-finally	{}		~	~

ظ ‹ẓ›:

	ط ‹ṭ›
word-finally	~

غ ‹ġ›:

	ك ‹k›	گ ‹ǵ›
word-initially	~	~
word-medially	~	~

ڤ ‹ṿ›:

	ف ‹f›
word-initially	~
word-medially	~
word-finally	~

ف ‹f›:

	ڤ ‹ṿ›	ق ‹q›	ڤ ‹v›
word-initially	~	{}	+
word-medially	~	~	+
word-finally	~	{}	

ق ‹q›:

	ف ‹f›	ڤ ‹v›	ك ‹k›	گ ‹ǵ›	ڮ ‹ĝ›
word-initially	{}	+	+	+	+
word-medially	~	+	+	+	+
word-finally	{}	+	+	+	+

ڤ ‹v›:

	ف ‹f›	ق ‹q›	ك ‹k›	گ ‹ǵ›	ڮ ‹ĝ›
word-initially	+	+	+	+	+
word-medially	+	+	+	+	+
word-finally		+	{}	+	+

ك ‹k›:

	ج ‹ǧ›	غ ‹ġ›	ق ‹q›	ڤ ‹v›
word-initially	~	~	+	+
word-medially	{}	~	+	+
word-finally			+	{}

گ ‹ǵ›:

	ج ‹ǧ›	غ ‹ġ›	ق ‹q›	ڤ ‹v›
word-initially	+	~	+	+
word-medially	+	~	+	+
word-finally			+	+

ڤ ‹ĝ›:

	ج ‹ğ›	ق ‹q›	ڤ ‹v›
word-initially	~	+	+
word-medially	~	+	+
word-finally		+	+

م ‹m›:

	ن ‹n›
word-medially	~

ن ‹n›:

	ا ‹a›	م ‹m›
word-medially		~
word-finally	~	

ه ‹h›:

	ا ‹a›	ة ‹ħ›	و ‹w›	ى ‹á›
word-finally	+	+	+	~

ة ‹ħ›:

	ا ‹a›	ت ‹t›	ه ‹h›	ى ‹á›
word-finally	+	+	+	+

و ‹w›:

	أ ‹å›	ه ‹h›	ۋ ‹ẘ›
word-initially	+		
word-medially			~
word-finally		+	

ۋ ‹ẘ›:

	أ ‹å›	و ‹w›
word-medially	~	~

ى ‹á›:

	ا ‹a›	ه ‹h›	ة ‹ħ›
word-finally	+	~	+

ي ‹y›:

	ا ‹a›	إ ‹ạ›	ئ ‹ỿ›
word-initially	+	+	
word-medially	+	+	+
word-finally	~		~

ئ ‹ỿ›:

	أ ‹å›	ي ‹y›	ء ‹o›
word-medially	~	+	~
word-finally		~	

ء ‹o›:	ئ ‹ỷ›
word-medially	~

The most general observation conerning this type of variation is that it is the most pervasive one. It affects 32 graphs out of a total of 42, i.e. more than 75 per cent. It is also responsible for much of the irregularity of MA graphy: It is here where marginal variations usually occur and expected forms abound. Four areas of a dense grid of restricted variation, due to the intersection of various spelling principles, can be observed:

(1) ا ‹a›, أ ‹å›, آ ‹ã› and إ ‹ą›, i.e. graphs with their *rasm*s based on the *ʔalif*, used word-initially for marking word-initial vowels or acting as pseudo-prothesis signalling clusters and pseudo-clusters.

(2) ا ‹a›, أ ‹å›, ه ‹h›, ة ‹ħ› and ى ‹á›, used word-finally to mark word-final *a*.

(3) ج ‹ǧ›, غ ‹ġ›, ق ‹q›, ڤ ‹v›, ك ‹k›, گ ‹ģ› and ڮ ‹ĝ› used to mark *g*.

(4) ا ‹a›, أ ‹å›, و ‹w›, ؤ ‹ẘ›, ي ‹y›, ئ ‹ỷ›, ء ‹o› marking *ʔ* or sounds corresponding to it in MA or merely copying the donor graphy.

The grid of relations between graphs bound by restricted variation is considerably less ordered in comparison with asymmetrical and free variations. Unlike these two variations, restricted variation does not presuppose transitivity: For instance, the fact that ق ‹q› Rv ك ‹k› and ك ‹k› Rv غ ‹ġ› does not necessarily mean that ق ‹q› Rv غ ‹ġ›. Many graphs bound by restricted variation also occur in pairs of graphic words in which they differentiate meanings context-independently, e.g. ق ‹q› and ك ‹k›. In many cases, variation is secondary while opposition is the primary relation. This harmonizes with the assumption made in Sect. III.2.2.2. that two graphs bound by restricted variation are not necessarily classified into one grapheme.

Graphs that have only restricted variants, i.e. no free or asymmetrical variants, e.g. س ‹s›, ص ‹ṣ› or م ‹m›, are placed in separate graphemes, with no allographs.

5. **Pseudo-variation**. Below, a list of pairs bound by pseudo-variation which have been identified in the corpus is given.

1. Intra-variety pseudo-variation.

ا ‹a› Pv ت ‹t›, word-initially, word-medially
ا ‹a› Pv ل ‹l›, word-initially, word-medially
ا ‹a› Pv ن ‹n›, word-initially
أ ‹å› Pv ت ‹t›, word-initially
أ ‹å› Pv ل ‹l›, word-initially
أ ‹å› Pv ن ‹n›, word-initially

2. Inter-variety pseudo-variation.
ا ‹a› Pv ل ‹l›, word-medially
أ ‹å› Pv ل ‹l›, word-medially
آ ‹ã› Pv و ‹w›, word-medially
ه ‹h› Pv ي ‹y›, word-finally

Intra-variety pseudo-variation manifests itself mostly in graphic words in which ا ‹a› or أ ‹å› are used to signal clusters and pseudo-clusters word-initially as pseudo-prothesis or to signal its remnant word-medially, as opposed to the graphs ت ‹t›, ل ‹l› and ن ‹n›, which are morphological graphies of the definite article, preposition *l(ə)-* 'to, for' or verbal affixes *t-* and *n-*. As for inter-variety pseudo-variation, it results, in most cases, from using ا ‹a›, أ ‹å›, آ ‹ã› and ه ‹h› as donor form-oriented graphies, as opposed to the self-oriented phonetic graphies ل ‹l› for *l*, و ‹w› for *ŭ* and ي ‹y› for *i*.

The fact that two graphs are bound by pseudo-variation does not entail classifying them into one grapheme.

6. **Graphemes**. Following the proposed criterion, an inventory of 34 MA graphemes has been arrived at, listed below (a basic allograph is given first, non-basic allographs are given after the colon):

ا ‹a›: إ ‹ą›	س ‹s›	ك ‹k›: گ ‹ģ›, ڪ ‹ĝ›
أ ‹å›	ش ‹š›	ل ‹l›
آ ‹ã›	ص ‹ṣ›	م ‹m›
ب ‹b›: پ ‹p›	ض ‹ḍ›: ظ ‹ẓ›	ن ‹n›
ت ‹t›: ث ‹t̲›	ط ‹ṭ›	ه ‹h›
ج ‹ǧ›	ع ‹ʕ›	ة ‹ħ›
ح ‹ḥ›	غ ‹ġ›	و ‹w›
خ ‹x›,	ڤ ‹ḟ›	ۏ ‹ẘ›
د ‹d›: ذ ‹d̲›	ف ‹f›	ى ‹á›
ر ‹r›	ق ‹q›	ي ‹y›
ز ‹z›	ڤ ‹v›	ئ ‹ŷ›
		ء ‹o›

Contrary to our initial expectations, graphs which are members of each of the following pairs: ا ‹a› ~ أ ‹å›, ا ‹a› ~ آ ‹ã›, ي ‹y› ~ ئ ‹ŷ›, ه ‹h› ~ ة ‹ħ›, ف ‹f› ~ ڤ ‹v› and ى ‹á› ~ ي ‹y› have been classified into different graphemes. The reasons for this are discussed in detail in respective places. As far as أ ‹å›, آ ‹ã› and ئ ‹ŷ› are con-

cerned, these graphs have turned out not only to function as donor form-oriented graphies for vowels but also to mark *ʔ*, a sound counted amoung MA phonemes (cf. Durand 2004: 71), and thus perform distinctive functions. The graphs ه ‹h› ~ ة ‹ħ› could not be classified into one grapheme because of their, albeit marginal, distinctive functions regarding the construct state. The fact that both ف ‹f› and ڤ ‹v› are used to mark *v* could suggest that they are allographs; however, they are not always substitutable: The former marks *q* in the local Maghrebi graphy, while the latter at times mark *g* – although it is unlikely for them to appear in a single text in these functions. Finally, the pair ى ‹á› ~ ي ‹y› shows free variation but this is manifested in too marginal a measure on the level of actual graphic words (in contradistinction to written EA, in which they would be most probably classified into one grapheme with similar criteria – an interesting difference between the graphies of the two dialects).

In contrast to the complex and intricate grid of relations of restricted variation, those of free and asymmetrical variations are simple and transparent. This results in a simple system of graphemes, with only six which have non-basic variants, out of the total 34 graphemes. Compared with the number of graphs that show restricted variation, this number is very moderate. Being unable to differentiate meanings and pronunciations, these non-basic variants are in a way superfluous units of the graphic system and questions can be asked about the reasons for their use. Two can be identified. One is that some non-basic allographs are phonetically less ambiguous than than their respective basic allographs. This concerns graphs used as a result of adaptive phonetic graphy: the exogenous graphs پ ‹p›, گ ‹ǵ› and ڴ ‹ĝ› marking *p* and *g* univocally (in contrast to the phonetically ambiguous ب ‹b› and ك ‹k›), as well as one case of an endogenous graph: إ ‹ạ›, which univocally marks *i* (in contrast to the phonetically ambiguous ا ‹a›). The other reason for the existence of non-basic allographs is that the three graphs ث ‹t̲›, ذ ‹d̲› and ظ ‹ẓ› are used as donor form-oriented graphies, i.e. they imitate the graphs marking interdental fricatives in SA (in contrast to ت ‹t›, د ‹d› and ض ‹ḍ›, used as self-oriented phonetic graphies). In this way, two pairs of opposing principles are reflected in the system of graphemes: that of economy vs. the phonetic principle on the one hand, and the principle of self-orientation and that of donor form-orientation, on the other.

Conclusion

Variation appears to be the most pervasive and at the same time one of the most attractive traits of MA graphy from a descriptive linguist's point of view. At the time the Quran was compiled, there was variation in the spelling of particular words as well (al-Ḥasan 2003: 362). The holy book of Islam is obviously utterly different in terms of content, function and status from the works of dialectal literature, which, being mostly of very wordly character, tend to concern trifling matters. However, in both cases, we are dealing with one of the spoken varieties of Arabic acquiring a written form in a spontaneous and natural process driven by insiders, not imposed from the outside, which is characteristic of so many dialects and languages used all over the world throughout history. This does not have to mean that one day standardized orthography will necessarily be created for MA. Should this happen, it cannot be guaranteed either that such a standardization will completely remove variation from texts written in this variety of Arabic, although it will surely diminish it, as was the case with SA.

At the present stage, with only five graphs that have no variants whatever, the vast domain of variation is used extensively and creatively by the spellers, who rely on the different spelling principles: To some extent, they couple the graphy of MA with the donor graphy, by borrowing its forms and principles, but also rebel against it. In many respects, they use self-oriented graphy, reflecting MA pronunciation and morphology, without forgetting, at times, the historical development of MA. However, it would be unjustified to claim that in writing MA anything goes. In fact, the main objective of the spellers, apart from striving for individuality and independence, is, at the end of the day, to make themselves understood. The intelligibility of MA written texts seems to be ensured by two factors. First, graphs which never vary form a core of stability. But the second one seems to be more important: The domain of variation is governed by principles, which this study has identified, at least partly, as far as the qualitative type is concerned. These principles regulating what and where can or cannot vary seem to be internalized by the users of written MA, even though they are never verbalized. Summing up, the macrosystem is fluctuating but it works because its users do not have problems recognizing and coping with variational heterographs.

The establishing of what is qualitatively variable and what is not has made it possible to propose a list of conditions for qualitative invariants on the one hand. On the other hand, it enables us to predict, at least to a certain extent, the existence

of expected forms, i.e. forms that have not been recorded in the corpus but most probably are used and most probably can be identified in an extended corpus. Positing expected forms should be, however, undertaken with caution because of two factors. One is the conditionality of some graphies, i.e. the dependence of a graphy on another. The other one is that some expected forms, predictable on the grounds by analogy, have not been recorded despite the high frequency of the use of the words they represent, most probably because of the weight of the orthography of the donor language (e.g. the case of *ḍar* 'house' written always دار ‹dar›, i.e. in the donor form-oriented manner, never phonetically ضار ‹ḍar›).

On the basis of this study, some new paths for investigation can be indicated. First of all, it should be followed by an examination of quantitative and linear variational heterographies, two important fields to explore. Such a study should result in establishing principles governing situations in which linear and quantitative variations are allowed and, on the other hand, identifying conditions for linear and quantitative invariants. Of special interest are interdependencies between the three types of heterography, which can be analyzed in the light of conditioned graphy. At this point, it seems doubtful that vocalization signs (secondary graphs) should be included into the analysis and examined from the perspective of quantitative variation. Even though they are at times co-functional with primary graphs (notably, the *šadda*), their irregular use and their virtually infinite quantitative variation (contrasting with their very limited qualitative variation) would make the description infinitely complex and hardly illuminative as far as the graphy of MA is concerned. Finally, since ambiguous forms can be observed within each type of heterography, it seems worthwhile to examine which type of ambiguous heterography is most frequently responsible for the context-dependent character of MA graphy and which one only exceptionally. Another question to answer is to what extent these three types of ambiguity can co-occur, i.e. how frequent is the use of graphic words which show linear, quantitative and qualitative ambiguity simultaneously.

Second, the corpus can be expanded. The examination of a larger quantity of texts will yield different findings, more accurately reflecting the reality of MA graphy. This would make it possible to verify whether what has been referred to as expected forms can be recorded or not. The expansion of the corpus can be done in three dimensions. First, a greater number of pages of the sources included in the present corpus could be examined. This, however, does not seem to be a fruitful line of enquiry since the samples of the sources included in the present corpus seem to be large enough for us to believe that not much new data will be recorded in their remaining parts (although some missing forms could be found there). The second manner of enlarging the corpus is to include other literary sources in it. It is desirable insofar as other authors are likely to use graphies not recorded

in the present corpus and, hence, not analyzed in the study. What is particularly missing in the present corpus are such graphic phenomena as the local Maghrebi graphy for *f* and *q*, practically non-existent there. It would be interesting to know if these graphs are ever used in the functions of marking the dialectal sounds *v* and *g*, which would be a situation of combining local classical graphy with local dialectal pronunciation. Adding new sources to the corpus could also be a helpful means of determining the formal features of some graphs, such as ڴ ‹ĝ› or ڤ ‹v›, over whose exact shape used in the non-neutralized position there is a question mark. Finally, the issue of marking *ʔ* in MA, for which the present corpus reveals a relatively strong variation, despite few occurrences, could be better elucidated when more words containing this sound could be examined. Enlarging the corpus by adding new titles to it would also be useful for a more thorough examination of the phenomenon of conditioned graphy: It could be particularly interesting to see in detail whether the donor-oriented or self-oriented graphy of element *X* entails a homogeneous graphy of element *Y* or not. The third way in which the corpus can be enlarged consists in including other text genres, such as advertising and journalistic. It can be expected that such texts can be equally well examined with the use of the proposed theoretical apparatus. Including them in the corpus will enable us to investigate to what extent graphies used in literary texts are different from those used by the spellers of advertisments which can be seen in Moroccan streets and magazines and by the spellers in the media. In particular, it should be examined whether the graphies used in the former genre are more varied or more complicated, due to spellers aspiring to some kind of graphic individualism, while that of the latter genres, being aimed at a larger audience, are rather simpler. A separate category of texts are those used in semi-private online communication (blogs, posts in social media, forums, comments, etc.). They tend to be considerably less carefully edited and their analysis and comparison with literary texts, written with the purpose of being printed, should be carried out with this in mind.

Third, the analysis of the present corpus, or an expanded one, could be examined with the use of statistical methods. For the time being, such an approach has been left for those whose domain it is, although it has been attempted in the course of the present work to characterize the identified graphies with respect to the frequency of their use, imprecise as such indications may appear. An exact measurement which type of graphies (morphological, phonetic, donor form-oriented etc.) are more frequent and which are marginal would be insightful. A rigorous quantitative examination of these phenomena could better reflect the fact, related to the macrosystem of the MA graphy being composed of a number of different, often internally incohesive microsystems, that some pairs of graphs can be bound by more than one relation in a single source or in various sources. In addition, the ratio of the use of particular spelling principles in the spelling macrosystem could

be calculated in order to determine the extent to which some principles are dominating, i.e. responsible for the greatest number of graphies.

Finally, a broader study could include comparison with the SA graphy and, more interestingly, with the graphy of other Arabic dialects. The best and nearly obvious choice would be EA, for which the corpus of modern written texts is vast, followed probably by Lebanese Arabic. Such a comparison, taking into consideration the issue of the homogeneity of text genres, would have to be preceded by conducting an analogous study of qualitative heterography in a given variety. The object of comparison should be not particular graphic words, as these are uncomparable due to the phonetic peculiarities of each variety, but principles underlying particular spellings. Interesting differences are bound to be identified.

References

Primary sources

The corpus

In parentheses, information is provided about the MA content in each analyzed book or its fragment (see the symbols *A*, *B* and *C* below), followed by the number of the examined pages and, if the pages are continuous, their range, after a colon. Asterisks mark books examined in their entirety.

A – pages written entirely or predominantly in MA.
B – pages containing paragraphs in MA.
C – pages written generally in SA but containing MA words, phrases or sentences.

The titles and the authors' names are transcribed in SA, if possible.

Aʕ *ʔAhl al-ʕatmāt*. ʕAbd al-Raḥmān Misḥat. [2005]. Beni Mellal (*C* 14).

BB *Bilād bəllarəž*. ʔAḥmad Luwayzī. 2011. Beirut: Dār al-ʔĀdāb (*B*, *C* 231*).

Bə. *Al-ʔaʕmāl al-kāmila. Al-masraḥiyyāt*, Part 1. ʕAbd al-Karīm Bərrəšid. 2009. [n.p.]: Manšūrāt Wizārat al-Ṯaqāfa (*A* 68: 22-89).

BT *Baqāyā min tīn al-ǧabal*. Al-Mīlūdī Šaġmūm. 2009. Latakia: Dār al-Ḥiwār (*B* 39).

Ḍʔ *Al-ḍarīḥ al-ʔāxar: sīra ḏātiyya*, ʕAbd al-Ġanī ʔAbū l-ʕAzm, 1996(?). Rabat: Muʔassasat al-Ġanī li-l-Našr (*B*, *C* 34).

Ḍa. *Al-ḍarīḥ: sīra ḏātiyya riwāʔiyya*, ʕAbd al-Ġanī ʔAbū l-ʕAzm. [n.d.]. Rabat (*B*, *C* 96: 10-105).

DB *Dumūʕ Bāxūs*. Muḥammad Manṣūr. 2010. [n.p.]: Éditions Mauja/al-Mawǧa (*B* 24).

ḌK *Al-ḍafādiʕ al-kəḥla: ʔinsāniyya min ʔarbaʕat ʔaḥdāṯ (masraḥiyya).* Muḥammad Šahramān. 1995. Marrakesh: Taǧhīzāt Makātib al-Ḥawz (*A* 31: 7-37).

FB *Fī al-baḥr ḏākiratī*. ʔAḥmad Ləmsiḥ. 2004. [n.p.] (*B* 38).

FL *Fās... law ʕādat ʔilayhi*. ʔAḥmad al-Madīnī. 2003. Rabat: Nādākūm li-l-Ṭibāʕa wa-l-Našr (*C* 25).

Fʕ *Lə-fšuš l-ʕəryan*. ʕAzīz r-Rəgragi. 2008(?). Rabat: Rabat Net Maroc (*A* 63*).

ḤB *Ḥəyyaḥat l-baša*. ʕAzīz r-Rəgragi. 2007(?). Rabat: Rabat Net Maroc (*A* 59*).

ḤD *Ḥallāq darb al-fuqarāʔ*. Yūsuf Fāḍil. 1991. Casablanca: ʕUyūn al-Maqālāt (*A* 20: 14-33).

ḤM *Ḥayawāt mutaǧāwira*. Muḥammad Barrāda. 2009. Casablanca: Našr al-Fanak (*A* 19: 74-92*).

LŠ *Layl al-šams*. ʕAbd al-Karīm Ǧuwayṭī. 1991. [n.p.]: Manšūrāt Ittiḥād Kuttāb al-Maġrib (*C* 12).

Mǧ. *Maǧnūn al-māʔ*. Dris Bəlmliḥ. 2004 (2010?). Casablanca (?): Manšūrāt Zāwiya (*B*, *C* 22).

MḤ *Al-maǧmūʕa al-ḥamrāʔ min al-nuṣūṣ al-masraḥiyya*. ʕAbd Allāh Šaqrūn. 2002. Casablanca: Maṭbaʕat al-Naǧāḥ al-Ǧadīda (*A* 59: 15-73).

Mi. *Mitru Muḥal.* Yūsuf Fāḍil. 2006. Casablanca: Našr al-Fanak (*B*, *C* 40).

Mm. *Mamlakat al-qaḥṭ.* ʕAbd al-ʔIlāh Buʕəsriya. 2006(?). [n.p.] (*A* 43: 111-153; pp. 127-153 contain poetry).

Mn. *Manām al-qiṭṭ: masraḥiyya fī ṯalāṯat fuṣūl*. Muḥammad Timəd. 1997 (*A* 87: 9-95).

Mu. *Mulat n-nuba*. 2006. ʕAzīz r-Rəgragi. Rabat: Rabat Net Maroc (*A* 60*).

MX *Al-maǧmūʕa al-xaḍrāʔ min al-nuṣūṣ al-masraḥiyya*. ʕAbd Allāh Šaqrūn. 2002. Casablanca: Maṭbaʕat al-Naǧāḥ al-Ǧadīda (*A* 210*).

QQ *Qayəd l-qīyyad (L-baša l-Glawi)*. ʕAbd al-ʔIlāh Bənhəddar. [n.d.]. Rabat (?): Saʕd al-Warzāzī li-l-Našr (*A* 49: 23-71).

Ṛḥ. *Ṛ-ṛḥil. Dəmʕa mṣafṛa*. Murād ʕAlamī. 2012. Rabat: Dār ʔAbī Raqrāq li-l-Ṭibāʕa wa-l-Našr (*A* 213*).

Ṣa. *Al-ʔaʕmāl al-kāmila*, Part 2. *Al-masraḥiyyāt*. Al-Miskīnī al-Ṣaġīr. 2007. [n.p.]: Manšūrāt Wizārat al-Ṯaqāfa (*A* 47: 208-254).

ṢX *Ṣabwa fī xarīf al-ʕumr*. Ḥasan Awrīd. 2006. Rabat: Manšūrāt Markaz Ṭāriq ibn Ziyād (*C* 10).

ṮM *3* [*Ṯalāṯ*] *masraḥiyyāt: Al-laǧna. Fitna. Ṯarṯara ʕalā ḍifāf ʔAbī Raqrāq*. Bašīr al-Qamarī. 2007. Kenitra: Al-Būkīlī li-l-Ṭibāʕa wa-l-Našr (*A* 66: 6-71).

ṮN *Ṯalāṯat nuṣūṣ masraḥiyya: N-nəqša. Al-mufattiš. Vūlpūn*. Al-Ṭayyib al-Ṣiddīqī. 2003. [n.p.]: Manšūrāt Wizārat al-Ṯaqāfa (*A* 54: 8-61).

WM *Wŭld Mimuna*. Muḥammad al-Ṭālibī. 2007. [n.p.]: Manšūrāt Wizārat al-Ṯaqāfa (*A* 31: 15-46).

XF *Xaṭṭ al-fazaʕ*. Dris Bəlmliḥ. 1998. Casablanca: ʔAfrīqiyā al-Šarq (*C* 16).

XM *Xamīl al-maḍāǧiʕ*. Al-Mīlūdī Šaġmūm. 1997(?). Mohammedia: Maṭbaʕat Faḍḍāla (*C* 93: 54-59, 66-152).

ʕD *ʕərs əd-didan*. Ləḥsən Ḥaddad. 2005. El Jadida: Ləḥsən Ḥaddad (*B*, *C* 44).

Other works in MA, not included in the corpus

Ami. *L-amir ṣ-ṣġiṛ*. ʕAbd al-Raḥīm Yūsī (Abderrahim Youssi). Translation of: Antoine de Saint-Exupéry, *Le Petit Prince*. 1st edn. 2009. Casablanca: Editions Aïni Bennaï. 2nd edn. 2014. Rabat: Kalimate.

Ḥaš. *Ḥašīš*. Yūsuf Fāḍil. 2000. (Casablanca): Našr al-Fanak.

Hhi. *Hhikayat aalamiya be ellougha elmeghribiya. Contes internationaux en langue marocaine*. Mourad Alami. 2009. Rabat: Bouregreg.

Tqə. *Tqərqib n-nab*. Yūsuf ʔAmīn al-ʕAlamī (Youssouf Amine Elalamy). 2006. Tangier: Xbaṛ bladna.

Works in SA

Ṯaʕ. *Al-ṯaʕlab allaḏī yaẓhar wa-yaxtafī*. Muḥammad Zafzāf. 2004. Baghdad: Manšūrāt al-Ǧamal – Köln: Al-Kamel Verlag.

ʔIṯ. *ʔIṯnā ʕašara raǧulan*. Yūsuf al-Sibāʕī. [n.d.]. Cairo: Maktabat Miṣr.

Taġ. *Taġrīdat al-baǧaʕa*. Makkāwī Saʕīd. 2009. Cairo: Al-Dār li-l-Našr wa-l-Tawzīʕ.

Works in EA

ʕAyz. *ʕAyza ʔatgawwiz*. Ġāda ʕAbd al-ʕĀl. 2009. Cairo: Dār al-Šurūq.

Secondary sources

Abu-Rabia, Salim & Taha, Haitham. 2013. "Reading in Arabic orthography: Characteristics, research findings, and assessment". In: Malatesha Joshi & Aaron 2013, 321-338.

Achour-Kallel, Myriam (ed.). 2015. *Le social par le langage – La parole au quotidien*. Tunis: IRMC – Paris: Karthala.

ʔAfā, ʕUmar & al-Maġrāwī, Muḥammad. 2007. *Al-xaṭṭ al-maġribī: Tārīx wa-wāqiʕ wa-ʔāfāq* [The Maghribi script: History, facts, perspectives]. Casablanca: Wizārat al-ʔAwqāf wa-l-Šuʔūn al-ʔIslāmiyya.

Aguadé, Jordi. 1996. "Notas acerca de los preverbios del imperfectivo en árabe dialectal magrebí". *Estudios de dialectología norteafricana y andalusí*", 1, 197-213.

Aguadé, Jordi. 2003. "Estudio descriptivo y comparativo de los fonemas del árabe dialectal marroquí". *Estudios de dialectología norteafricana y andalusí*, 7, 59-109.

Aguadé, Jordi. 2005. "*Darle al pico*: un 'bestiario' de Youssouf Amine Elalamy en árabe marroquí". *Estudios de dialectología norteafricana y andalusí*, 9, 245–265.

Aguadé, Jordi. 2006. "Writing dialect in Morocco". *Estudios de dialectología norteafricana y andalusí*, 10, 253-274.

Aguadé, Jordi, 2008. "Morocco". In: Versteegh 2006-2009, vol. 3, 287-297.

Aguadé, Jordi, 2010. "On vocalism in Moroccan Arabic dialects". In: Monferrer-Sala & Al Jallad 2010, 95-105.

Aguadé, Jordi. 2012. "Monarquía, dialecto e insolencia en Marruecos: el caso *Nichane*". In: Meouak *et al.* 2012, 441-464.

Aguadé, Jordi. 2013. "Des romans diglossiques: le cas de Youssef Fadel". In: Benítez Fernández *et al.* 2013, 207-220.

Aguadé, Jordi & Benyahia, Laila. 2005, *Diccionario árabe marroquí: Árabe marroquí-español, español-árabe marroquí*. Cádiz: Quorum Editores.

Al-Ani, Salman H. 2008. "Phonetics". In: Versteegh 2006-2009, vol. 3, 593-603.

Alosh, Mahdi. 2005. *Using Arabic: A guide to contemporary usage*. Cambridge: Cambridge University Press.

Al-Toma, Salih J. 1961. "The Arabic writing system and proposals for its reform". *Middle East Journal* 15, 4, 403-415.

ʔAmṯāl. 2001 = *Al-ʔamṯāl al-ʕāmmiyya fī al-Maġrib: Tadwīnuhā wa-tawẓīfuhā al-ʕilmī wa-l-bīdāġūǧī. Nadwat Laǧnat al-Turāṯ al-tābiʕa li-ʔAkādīmiyyat al-Mamlaka al-Maġribiyya bi-l-mušāraka maʕ al-Ǧamʕiyya al-Maġribiyya li-l-Turāṯ al-Luġawī* [Dialectal proverbs in Morocco: Their recording and scholarly and pedagogical use. Session of the Heritage Committee of the Academy of the Kingdom of Morocco with the participation of the Moroccan Society for Linguistic Heritage]. Rabat: ʔAkādīmiyyat al-Mamlaka al-Maġribiyya.

ʔAsmar [al-], Rāǧī. 1988. *Al-marǧiʕ fī al-ʔimlāʔ* [Handbook of spelling]. Revised by ʔImīl Badīʕ YaʕqūbTripoli: Ǧrūs Bris.

Augst, Gerhard (ed.). 1986. *New trends in graphemics and orthography*. Berlin – New York: Walter de Gruyter.

Badawī, al-Saʕīd Muḥammad. 1973. *Mustawayāt al-ʕarabiyya al-muʕāṣira fī Miṣr: Baḥṯ fī ʕalāqat al-luġa bi-l-ḥaḍāra* [The levels of Modern Arabic in Egypt: A study of the relationship between language and culture]. Cairo: Dār al-Maʕārif.

Badawi, Elsaid; Carter, M.G. & Gully, Adrian. 2004. *Modern Written Arabic: A comprehensive grammar*. New York: Routledge.

Bańczerowski, Jerzy; Pogonowski, Jerzy & Zgółka, Tadeusz. 1982. *Wstęp do językoznawstwa* [Introduction to linguistics]. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza w Poznaniu.

Barontini, Alexandrine; Pereira, Christophe; Vicente, Ángeles & Ziamari, Karima (eds.). 2012. *Dynamiques langagières en Arabophonies: Variations, contacts, migrations et créations artistiques. Hommage offert à Dominique Caubet par ses élèves et collègues*. Zaragoza: Universidad de Zaragoza.

Bassiouney, Reem. 2009. *Arabic sociolinguistics*. Edinburgh: Edinburgh University Press.

Baudouin de Courtenay, Jan. 1901. "Wskazówki dla zapisujących materjały gwarowe na obszarze językowym polskim" [Indications for persons recording dialect material in the Polish-speaking area]. In: *Materyały* 1901, 115-139.

Bauer, Thomas. 1996. "Die schriftliche Sprache im Arabischen". In: Günther & Ludwig 1996, 1483-1490.

Beeston, A.F.L. 1968. *Written Arabic: An approach to the basic structures*. Cambridge: The University Press.

Beeston, A.F.L. 1970. *The Arabic language today*. London: Hutchinson University Library.

Beißwenger, Michael. 2001. *Chat-Kommunikation: Sprache, Interaktion, Sozialität & Identität in synchroner computervermittelter Kommunikation: Perspektiven auf ein interdisziplinäres Forschungsfeld*. Stuttgart: Ibidem.

Belnap, R. Kirk & Bishop, Brian. 2003. "Arabic personal correspondence: a window on change in progress?" *International Journal of the Sociology of Language*, 163, 9-25.

Benhallam, Abderrafi. 1980. *Syllable structure and rule types in Arabic*. Gainesville: University of Florida. Unpublished PhD dissertation.

Benhallam, Abderrafi. 1989/1990. "Moroccan Arabic syllable structure". *Langues et littératures* (Rabat), 8, 177-191.

Benítez, Montserrat. 2003. "Transcripción al árabe marroquí de mensajes de teléfono móvil". *Estudios de dialectología norteafricana y andalusí* 7, 153-163.

Benítez Fernández, Montserrat. 2010. *La política lingüística contemporánea de Marruecos: de la arabización a la aceptación del multilingüismo*. Zaragoza: Instituto de Estudios Islámicos y del Oriente Próximo.

Benítez Fernández, Montserrat. 2012a. "*TelQuel*: una fuente contemporánea para el estudio del árabe marroquí". In: Meouak *et al.* 2012, 403-417.

Benítez Fernández, Montserrat. 2012b. "*Al-ʔAmal*: Otro intento fallido de escribir en *dārīža* marroquí". In: Barontini *et al.* 2012, 379-391.

Benítez Fernández, Montserrat; Miller, Catherine; de Ruiter, Jan Jaap & Tamer, Youssef (eds.). 2013. *Évolution des pratiques et représentations langagières dans le Maroc du XXI^e siècle*, vol. 1. Paris: L'Harmattan.

Benjelloun, Saïd. 1998. *El-luġa dyáli: El-ʕarabiya el-maġribiya* [My language: Moroccan Arabic], vol. 2. Tilburg: Syntax Publishers.

Berjaoui, Naser. 2001. "Aspects of the Moroccan Arabic orthography with preliminary insights from the Moroccan computer-mediated communication". In: Beißwenger 2001, 431-465.

Blau, Joshua. 1970. *On pseudo-corrections in some Semitic languages*. Jerusalem: Israel Academy of Sciences and Humanities.

Boogert [van den], N[ico]. 1989. “Some notes on Maghribi script”. *Manuscripts of the Middle East* 4, 30-43.

Bosworth, C.E.; van Donzel, E.; Lewis, B. & Pellat, Ch. (eds.). 1986-2004. *The encyclopedia of Islam. New edition*. 13 vols. Leiden: Brill.

Boudlal, Abdelaziz. 2011. “On the special behavior of schwa in Moroccan Arabic prosody”. Paper presented at the 8th Old World Conference in Phonology, Marrakesh, 19-22 January 2011 (https://www.academia.edu/6420600/On_the_Special_Behavior_of_Schwa_in_Moroccan_Arabic, accessed 17 December 2016).

Brend, Ruth M. (ed.). 1992. *The eighteenth LACUS forum, 1991*. Lake Bluff: Linguistic Association of Canada and United States.

Brunot, Louis. 1931-1952. *Textes arabes de Rabat. I. Textes, transcription et traduction annotée. II. Glossaire*. Paris: Publications de l’Institut des Hautes Études Marocaines.

Buckwalter, Tim. 2004. “Issues in Arabic orthography and morphology analysis”. In: Farghaly & Megerdoomian 2004, 31-34.

Bußmann, Hadumod. 1990. *Lexikon der Sprachwissenschaft*. Stuttgart: Alfred Kröner Verlag.

Cachia, P.[ierre]. 2010. “Dialect in literature, modern”. In: Meisami & Starkey 2010, 190-191.

Cantineau, Jean. 1950. “Réflexions sur la phonologie de l’arabe marocain”. *Hespéris* 37, 193-207.

Carney, Edward. 1994. *A survey of English spelling*. London – New York: Routledge.

Caubet, Dominique. 1993. *L’arabe marocain*, vol. I. *Phonologie et morphosyntaxe*. Paris-Louvain: Éditions Peeters.

Caubet, Dominique. 1999. “Arabe maghrébin: passage à l’écrit et institutions”. *Faits de langues* (*Oral-Ecrit: Formes et théories*) 13, March 1999, 235-244.

Caubet, Dominique. 2004. “L’intrusion des téléphones portables et des ‘SMS’ dans l’arabe marocain en 2002-2003”. In: Caubet *et al.* 2004, 247-270.

Caubet, Dominique. 2008. “Moroccan Arabic”. In: Versteegh 2006-2008, vol. 3, 273-287.

Caubet, Dominique. 2012. “Apparition massive de la *darija* à l’écrit à partir de 2008-2009: sur le papier et sur la toile: quelle graphie? Quelles régularités?”. In: Meouak 2012, 377-402.

Caubet, Dominique. 2013. “Maroc 2011 – Messagerie instantanée sur l’internet marocain: *facebook*, *darija* et parlers jeunes”. In: Benítez Fernández *et al.* 2013, 63-87.

Caubet, Dominique. 2017. “Morocco: An informal passage to literacy in *dārija* (Moroccan Arabic)”. In: Høigilt & Mejdell 2017, 116-141.

Caubet, Dominique; Billiez, Jacqueline; Bulot, Thierry; Léglise, Isabelle & Miller, Catherine (eds.). 2004. *Parlers jeunes ici et là-bas: Pratiques et représentations*. Paris: L’Harmattan.

Colin, G.S. 1986. “Linguistic survey” (part VII of the article “al-Maghrib”). In: Bosworth *et al.* 1986-2004, vol. 5, 1203-1207.

Corriente, Federico. 2002. *Gramática árabe*. Barcelona: Herder.

Coulmas, Florian. 2003. *Writing systems: An introduction to their linguistic analysis*. Cambridge: Cambridge University Press.

Crystal, David. 2008. *Dictionary of linguistics and phonetics*. Malden – Oxford: Blackwell.

Daḥdāḥ [al-], ʔAnṭuwān. 1989. *Muʕǧam qawāʕid al-luġa al-ʕarabiyya fī ǧadāwil wa-lawḥāt* [A dictionary of Arabic grammar in charts and pictures]. Beirut: Maktabat Lubnān.

Danecki, Janusz. 1989. *Wstęp do dialektologii języka arabskiego* [An introduction to the dialectology of Arabic]. Warszawa: Wydawnictwa Uniwersytetu Warszawskiego.

Danecki, Janusz. 1994. *Gramatyka języka arabskiego* [A grammar of Arabic]. Warszawa: Wydawnictwo Akademickie Dialog.

Dānī [al-], ʔAbū ʕAmr ʕUṯmān b. Saʕīd. 1997. *Al-muḥkam fī naqṭ al-maṣāḥif* [The perfect book on the pointing of the Qurans]. Ed. ʕIzza Ḥasan. Beirut: Dār al-Fikr al-Muʕāṣir – Damascus: Dār al-Fikr.

Daniels, Peter T. 1992. "Is a structural graphemics possible?". In: Brend 1991, 528-537.

Daniels, Peter T. 1995. "Reply to Herrick". In: Powell 1995, 425-431.

Daniels, Peter T. 2013. "The Arabic writing system". In: Owens 2013, 412-432.

Daniels, Peter T. 2014. "The type and spread of Arabic script". In: Mumin & Versteegh 2014, 25-39.

Diem, Werner. 1976. "Some glimpses at the rise and early development of the Arabic orthography". *Orientalia*, 45, 251-261.

Diem, Werner. 1979-1983. "Untersuchungen zur frühen Geschichte der arabischen Orthographie". *Orientalia*, 48, 207-257; 49, 67-106; 50, 332-383; 52, 357-404.

Diem, Werner. 1982. "Die Entwicklung der arabischen Orthographie". In: Fischer 1982, 184-190.

Diem, Werner. 2006. *Hochsprache und Dialekt im Arabischen: Untersuchungen zur heutigen arabischen Zweisprachigkeit*. Wiesbaden: Harrassowitz Verlag.

Durand, Olivier. 2004. *L'arabo del Marocco: Elementi di dialetto standard e mediano*. Roma: Università degli Studi La Sapienza.

Durand, Olivier; Langone, Angela Daiana & Mion, Giuliano (eds.). 2014. *Alf lahğa wa lahğa. Proceedings of the 9th Aida Conference.* Wien – Berlin: LIT Verlag.

Dürscheid, Christa. 2006. *Einführung in die Schriftlinguistik*. Göttingen: Vandenhoeck & Ruprecht.

Eid, Mushira & McCarthy, John (eds.). 1990. *Perspectives on Arabic linguistics II. Papers from the Annual Symposium on Arabic Linguistics*. Amsterdam – Philadelphia: John Benjamins Publishing.

El-Hassan, S.A. 1978. "Variation in the demonstrative system in Educated Spoken Arabic". *Archivum Linguisticum* 9, 1, 32-57.

Elinson, Alexander E. 2013. "*Dārija* and changing writing practices in Morocco". *International Journal of Middle East Studies* 45, 4, 715-730.

Elinson, Alexander E. 2017. "Writing oral and literary culture: The case of the contemporary Moroccan *zajal*". In: Høigilt & Mejdell 2017, 190-211.

Ennaji, Moha. 2005. *Multilingualism, cultural identity, and education in Morocco*. New York: Springer.

Ennaji, Moha. 2013. "Multilingualism in Morocco and the linguistic features of the Casablanca variety". In: Singleton *et al.* 2013, 231-246.

Farghaly, Ali & Megerdoomian, Karine (eds.). 2004. *Proceedings of the Workshop on Computational Approaches to Arabic Script-based Languages*. Stroudsburg: Association for Computational Linguistics.

Fāsī [al-], Muḥammad. 1986. *Maʕlamat al-malḥūn* [The encyclopedia of *malḥūn* poetry]. Part 1, vol. 1. Rabat: ʔAkādīmiyyat al-Mamlaka al-Maġribiyya.

Fazzini, Elisabetta (ed.). 2014. *Culture del Mediterraneo: Radici, contatti, dinamiche*. Milano: LED.

Ferguson, Charles A. 1959. "Diglossia". *Word* 15, 325-340.

Ferrando, Ignacio. 2012a. "Apuntes sobre el uso del dialecto en la narrativa marroquí moderna". In: Meouak *et al.* 2012, 349-358.

Ferrando, Ignacio. 2012b. "El árabe marroquí en la novela del siglo XXI. El caso de Salem Himmich y su obra فتنة الرؤوس والنسوة". In: Barontini *et al.* 2012, 415-426.

Fischer, Wolfdietrich (ed.). 1982. *Grundriß der arabischen Philologie*, vol. 1. Wiesbaden: Ludwig Reichert Verlag.

Gacek, Adam. 2001. *The Arabic manuscript tradition: A glossary of technical terms and bibliography*. Leiden – Boston – Köln: Brill.
Gacek, Adam. 2008. "Maġribī". In: Versteegh 2006-2009, vol. 3, 110-113.
Gacek, Adam. 2009. *Arabic manuscripts: A vademecum for readers*. Leiden – Boston: Brill.
Ġalāyīnī [al-], Muṣṭafā. 2002. *Ǧāmiʕ al-durūs al-ʕarabiyya: Mawsūʕa fī ṯalāṯat ʔaǧzāʔ*. [A compendium of Arabic lessons: An encyclopedia in three parts]. Sidon – Beirut: Al-Maktaba al-ʕAṣriyya. [1st edn 1912.]
Goldenberg, Gideon. 2013. *Semitic languages: Features, structures, relations, processes*. Oxford: Oxford University Press.
"Grapheme, *n*." 2018. *OED Online*. Oxford University Press, www.oed.com/view/Entry/80826 (accessed 18 April 2018).
Grigore, George & Bițună, Gabriel (eds.). 2016. *Arabic varieties – Far and wide. Proceedings of the 11th International Conference of AIDA – Bucharest, 2015*. București: Editura Universității din București.
Grotzfeld, Heinz. 1982. "Neuarabische Dialekte als Sprache der Literatur". In: Fischer 1982, 119-124.
Gruendler, Beatrice. 2006. "Arabic alphabet: Origin". In: Versteegh 2006-2009, vol. 1, 148-165.
Günther, Hartmut. 1988. *Schriftliche Sprache: Strukturen geschriebener Wörter und ihre Verarbeitung beim Lesen*. Tübingen: Max Niemeyer Verlag.
Günther, Hartmut & Ludwig, Otto (eds.). 1996. *Schrift und Schriftlichkeit: Ein interdisziplinäres Handbuch internationaler Forschung / Writing and its use: An interdisciplinary handbook of international research*, vol. 2. Berlin – New York: Walter de Gruyter.
Hammarström, Göran. 1966. *Linguistische Einheiten im Rahmen der modernen Sprachwissenschaft*. Berlin – Heidelberg – New York: Springer-Verlag.
Hammarström, Göran. 1976. *Linguistic units and items*. Berlin – Heidelberg – New York: Springer-Verlag.
Hammarström, Göran. 1981. "Type and typeme, graph and grapheme". In: Ruszkiewicz 1981a, 89-99 [Reprinted from *Studia Neophilologica* 1964, 36, 332-340.]
Hamzaoui, Rached. 1975. *L'Académie de langue arabe du Caire: Histoire et œuvre*. Tunis: Publications de l'Université de Tunis.
Harrell, Richard S. 1962. *A short reference grammar of Moroccan Arabic*. Washington, D.C.: Georgetown University Press.
Harrell, Richard S. (ed.). 1966. *A dictionary of Moroccan Arabic: Moroccan-English*. Washington, D.C.: Georgetown University Press.
Hārūn, ʕAbd al-Salām Muḥammad. 1993. *Qawāʕid al-ʔimlāʔ* [Spelling rules]. Cairo: Maktabat al-ʔAnǧlū al-Miṣriyya.
Ḥasan [al-], Ṣāliḥ b. Ibrāhīm. 2003. *Al-kitāba al-ʕarabiyya min al-nuqūš ʔilā al-kitāb al-maxṭūṭ* [The Arabic script from inscriptions to manuscripts]. Riyadh: Dār al-Fayṣal al-Ṯaqāfiyya.
Ḥasanayn, ʔAḥmad Ṭāhir & Šaḥāta, Ḥasan. 1998. *Qawāʕid al-ʔimlāʔ al-ʕarabī bayn al-naẓariyya wa-l-taṭbīq* [Arabic spelling rules between theory and practice]. Cairo: Maktabat al-Dār al-ʕArabiyya li-l-Kitāb.
Herrero Muñoz-Cobo, Barbara; Pérez Cañada, Luis Miguel; Aragón Huerta, Mercedes & Moscoso García, Francisco (eds.) 2011. *IV Congreso Árabe Marroqui: más allá de la oralidad. Toledo, 23 y 24 abril de 2010*. [Almería]: Universidad de Almería.
Herrick, Earl M. 1995a. "Of course a structural graphemics is possible!". In: Powell 1995, 413-424.

Herrick, Earl M. 1995b. “Reply to Daniels’ reply”. In: Powell 1995, 432-40.
“Heterography, *n*.” 2018. *OED Online*. Oxford University Press, www.oed.com/view/Entry/86468 (accessed 18 April 2018).
Høigilt, Jacob & Mejdell, Gunvor (eds.). 2017. *The politics of written language in the Arab world: Writing change*. Leiden – Boston: Brill.
Holes, Clive. 2004. *Modern Arabic: Structures, functions and varieties*. Washington: Georgetown University Press.
Hoogland, Jan. 1983. *Het Marokkaans Arabisch als schriftelijk medium*. Utrecht: unpubl. MA thesis.
Hoogland, Jan. 1996. *Marokkaans Arabisch: Een cursus voor zelfstudie en klassikaal gebruik*. Amsterdam: Bulaaq.
Hoogland, Jan. 2013a. “Towards a standardized orthography of Moroccan Arabic based on best practices and common ground among a selection of authors”. In: Santillán Grimm *et al.* 2013, 59-76.
Hoogland, Jan. 2013b. “L’arabe marocain, langue écrite”. In: Benítez Fernández *et al.* 2013, 175-188.
Hoogland, Jan. 2018 [to appear]. “Darija in the Moroccan press: The case of the magazine *Nichane*”. *Sociolinguistic Studies*.
Houdas, O. 1886. “Essai sur l’écriture maghrebine”. In: *Nouveaux mélanges* 1886, 83-112.
Ibn Durustawayh, ʔAbū Muḥammad ʕAbd Allāh b. Ǧaʕfar b. Muḥammad. 1921. *Kitāb al-kuttāb* [The book of scribes]. Ed. Luwīs Šayxū (Louis Cheikho). Beirut: Maṭbaʕat al-ʔĀbāʔ al-Yasūʕiyyīn.
Kästner, Hartmut. 1981. *Phonetik und Phonologie des Modernen Hocharabisch*. Leipzig: VEB Verlag Enzyklopädie.
Kaye, Alan S. 1994. “Formal vs. informal in Arabic: Diglossia, triglossia, tetraglossia, etc., polyglossia – multiglossia viewed as a continuum”. *Zeitschrift für Arabische Linguistik*, 27, 47-66.
Kew, Jonathan. 2005. “Notes on some Unicode Arabic characters: recommendations for usage”. (Draft 2 – April 21, 2005). http://scripts.sil.org/cms/scripts/page.php?site_id=nrsi&item_id=PCUnicodeDocs#272ccb85 (accessed 31 January 2018).
Kjamilev, S.X. 1968. *Marokkanskij dialekt arabskogo jazyka*. Moskva: Izdatel’stvo Nauka.
Kohrt, Manfred. 1985. *Problemgeschichte des Graphembegriffs und des frühen Phonembegriffs*. Tübingen: Max Niemeyer Verlag.
Kohrt, Manfred. 1986. “The term ‘grapheme’ in the history and theory of linguistics”. In: Augst 1986, 80-96.
Król, Iwona. 2009. *Pismo i ortografia arabska* [The Arabic script and orthography]. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego.
Lafkioui, Mena & Merolla, Daniela (eds.). 2008. *Oralité et nouvelles dimensions de l’Oralité: Intersections théoriques et comparaisons des matériaux dans les études africaines*. Paris: Inalco – Publications Langues’O.
Langone, Angela Daiana. 2003. “*Ḥbār blādna*. Une expérience journalistique en arabe dialectal marocain”. *Estudios de dialectología norteafricana y andalusí*, 7, 143-151.
Langone, Angela Daiana. 2006. “Jeux linguistiques et nouveaux style dans la masrahiyya «ənnəqša», Le déclic, écrite en dialectal marocain par Ṭ-Ṭiyyəb əṣ-Ṣaddīqi. In: Mejri 2006, 243-261.
Langone, Angela Daiana. 2008. “L’arabe dialectal passe à l’écrit. Quelques remarques sur le Maroc”. In: Lafkioui & Merolla 2008, 51-65.

Laroui, Fouad. 2011. *Le drame linguistique marocain*. Léchelle: Zellige.
Lerchundi, Joseph. 1900. *Rudiments of the Arabic-vulgar of Morocco*. Transl. James Maciver MacLeod. Tangier: The Spanish Catholic Mission Press. [Original version: 1872. *Rudimentos del árabe vulgar que se habla en el Imperio de Marruecos, con numerosos ejercicios y temas aplicados a la teoría*. Madrid: M. Rivadeneyra.]
Malatesha Joshi, R. & Aaron, P.G. (eds.). 2013. *Handbook of orthography and literacy*. London – New York: Routledge.
Materyały. 1901 = *Materyały i Prace Komisyi Językowej Akademii Umiejętności w Krakowie*. Tom I, zeszyt I. Kraków: Akademia Umiejętności.
Meisami, Julie Scott & Starkey, Paul (eds.). 2010. *The Routledge encyclopedia of Arabic literature*. London – New York: Routledge.
Meiseles, Gustav. 1980. "Educated Spoken Arabic and the Arabic language continuum". *Archivum Linguisticum*, 11, 118-148.
Mejdell, Gunvor. 2014. "Strategic bivalency in written 'mixed style'? A reading of Ibrāhīm ʿĪsā in *al-Dustūr*". In: Durand *et al.* 2014, 273-278.
Mejri, Salah (ed.). 2006. *L'arabe dialectal: Enquêtes, descriptions, interprétations. Actes d'AIDA 6*. Tunis: Centre d'Études et de Recherches Économiques et Sociales.
Meouak, Mohamed; Sánchez, Pablo & Vicente, Angeles (eds.). 2012. *De los manuscritos medievales a internet: la presencia del árabe vernáculo en las fuentes escritas*. Zaragoza: Universidad de Zaragoza.
Meynet, Roland. 1971. *L'écriture arabe en question: Les projets de l'Académie de Langue Arabe du Caire de 1938 à 1968*. Beirut: Dar El-Machreq.
Michalski, Marcin. 2016. "Spelling Moroccan Arabic in Arabic script: The case of literary texts". In: Grigore & Biţună 2016, 385-394.
Michalski, Marcin. 2017. "Describing Written Moroccan Arabic: Some methodological issues". *Rocznik Orientalistyczny*, 70, 2, 232-244.
Midlāwī [al-], Muḥammad. 2001. *Naḥw tadwīn al-ʔādāb al-šafahiyya al-maġribiyya fī ʔiṭār taṭwīr al-ḥarf al-ʕarabī al-muwassaʕ* [Toward recording Moroccan oral literature within the framework of developing expanded Arabic script]. In: *ʔAmṯāl* 2001, 159-211.
Miller, Catherine. 2012. "Observations concernant la présence de l'arabe marocain dans la presse marocaine arabophone des années 2009-2010". In: Meouak *et al.* 2012, 419-440.
Miller, Catherine. 2015. "Des passeurs individuels au mouvement linguistique: itinéraires de quelques traducteurs au Maroc". In: Achour-Kallel 2015, 203-232.
Miller, Catherine. 2017. "Contemporary *dārija* writings in Morocco: Ideology and practices". In: Høigilt & Mejdell 2017, 90-115.
Mion, Giuliano. 2014. "Nuove versioni del «Petit Prince» in arabo maghrebino". In: Fazzini 2014, 187-198.
Monferrer-Sala, Juan Pedro & Al Jallad, Nader (eds.). 2010. *The Arabic language across the ages*. Wiesbaden: Reichert.
Moroccan Arabic. 2011 = *Moroccan Arabic: Moroccan Arabic Textbook. Al-dāriğa al-maġribiyya: Ka-nətkəlləm d-dariža l-maġribiya/Kitāb al-dāriğa al-maġribiyya*. Rabat: Peace Corps Morocco.
Moscoso García, Francisco. 2006. *Curso de árabe marroquí: Diálogos, gramática, ejercicios, glosario, bibliografía*. Cádiz: Servicio de Publicaciones de la Universidad de Cádiz.

Moscoso García, Francisco. 2009. “Comunidad lingüística marroquí en los foros y chats. Expresión escrita, ¿norma o anarquía?”. *Al-Andalus Magreb*, 16, 209-226.

Mumin, Meikal & Versteegh, Kees (eds.). 2014. *The Arabic script in Africa: Studies in the use of a writing system*. Leiden – Boston: Brill.

Nouveaux mélanges. 1886 = *Nouveaux mélanges orientaux: Mémoires, textes et traductions publiés par les professeurs de l’École Speciale des Langues Orientales Vivantes à l’occasion du Septième Congrès International des Orientalistes réuni à Vienne*. Paris: Ernest Leroux.

Otten, Roel & Hoogland, Jan. 1983. *Basiswoordenboek van het Marokkaans Arabisch: Marokkaans-Nederlands, Nederlands-Marokkaans*. Muiderberg: Coutinho.

Owens, Jonathan (ed.). 2013. *The Oxford handbook of Arabic linguistics*. Oxford etc.: Oxford University Press.

Parkinson, Dilworth. 1990. “Orthographic variation in Modern Standard Arabic: The case of the *hamza*”. In: Eid & McCarthy 1990, 269-295.

Pellat, Charles. 1991. “Al-Malḥūn”. In: Bosworth *et al.* 1986-2004, vol. 6, 247-257.

Pérez Cañada, Luiz Miguel & Salinitro, Anna. 2011. “La edición en árabe marroquí: creación y traducción”. In: Herrero Muñoz-Cobo *et al.* 2011, 289-320.

Polański, Kazimierz (ed.). 2003a. *Encyklopedia językoznawstwa ogólnego* [Encyclopedia of general linguistics]. Wrocław: Zakład Narodowy im. Ossolińskich.

Polański, Kazimierz. 2003b. “Analogia w języku” [Analogy in language]. In: Polański 2003a, 46-47.

Polański, Kazimierz. 2003c. “Allograf” [Allograph]. In: Polański 2003a, 38-39.

Polański, Kazimierz (ed.). 2016a. *Wielki słownik ortograficzny PWN z zasadami pisowni i interpunkcji* [The PWN’s great orthographic dictionary with spelling and punctuation rules]. Warszawa: Wydawnictwo Naukowe PWN SA.

Polański, Kazimierz. 2016b. “Zasady pisowni i interpunkcji” [The spelling and punctuation rules]. In: Polański 2016a, 11-154.

Powell, Mava Jo (ed.). 1995. *The twenty-first LACUS forum, 1994*. Chapel Hill: Linguistic Association of Canada and the United States.

Prémare [de], Alfred-Louis. 1993-1999. *Dictionnaire arabe-français*. 12 vols. Paris: L’Harmattan.

Qabbiš, ʔAḥmad. 1984. *Al-ʔimlāʔ al-ʕarabī: Našʔatuhuwaqawāʕiduhuwamufradātuhuwatamrīnātuhu* [Arabic orthography: Its origin, rules, terminology and exercises]. Damascus – Beirut: Dār al-Rašīd.

Qalqašandī [al-], ʔAbū l-ʕAbbās ʔAḥmad b. ʕAlī. [n.d.]. *Ṣubḥ al-ʔaʕšā fī ṣināʕat al-ʔinšā* [Dawn for the night-blind: On the art of composition]. 14 vols. Cairo: Wizārat al-Ṯaqāfa wa-l-ʔIršād al-Qawmī [Reprint of al-Ṭabʕa al-ʔAmīriyya, 1913-1918].

Rogers, Henry. 2005. Writing systems: A linguistic approach. Malden – Oxford: Blackwell Publishing.

Rosenbaum, Gabriel M. 2004. “Egyptian Arabic as a written language”. *Jerusalem Studies in Arabic and Islam*, 29, 281-340.

Ruszkiewicz, Piotr (ed.). 1981a. *Graphophonemics: A book of readings*. Katowice: Uniwersytet Śląski.

Ruszkiewicz, Piotr. 1981b. “Jan Baudouin de Courtenay’s theory of the grapheme”. In: Ruszkiewicz 1981a, 20-34. [Reprinted from *Acta Philologica* 1978, 7, 117-135.]

Saiegh-Haddad, Elinor & Henkin-Roitfarb, Roni. 2014. “The structure of Arabic language and orthography”. In: Saiegh-Haddad & Malatesha Joshi 2014, 3-28.

Saiegh-Haddad, Elinor & Malatesha Joshi, R. (eds.). 2014. *Handbook of Arabic literacy: Insights and perspectives*. Dordrecht – Heidelberg – New York – London: Springer.

Sánchez, Pablo. 2014. *El árabe vernáculo de Marrakech: Análisis lingüístico de un corpus representativo*. Zaragoza: Prensas de la Universidad de Zaragoza.

Santillán Grimm, Paula; Pérez Cañada, Luis Miguel & Moscoso García, Francisco (eds.). 2013. *Árabe marroquí: de la oralidad a la enseñanza*. Cuenca: Ediciones de la Universidad de Castilla-La Mancha.

Sawaie, Mohammed. 2007. "Language academies". In: Versteegh *et al*. 2006-2009, vol. 2, 634-642.

Singleton, David; Fishman, Joshua A.; Aronin, Larissa & Ó Laoire, Muiris (eds.). 2013. *Current multilingualism: A new linguistic dispensation*. Boston – Berlin: Walter de Gruyter.

Srhir, Adil Moustaoui. 2016. *Sociolinguistics of Moroccan Arabic: New topics*. Frankfurt am Main: Peter Lang.

Ṭabbāʕ [al-], ʕUmar Fārūq. 1993. *Al-wasīṭ fī qawāʕid al-ʔimlāʔ wa-l-ʔinšāʔ* [The facilitator of spelling and writing]. Beirut: Maktabat al-Maʕārif.

Versteegh, Kees (ed.). 2006-2009. *Encyclopedia of Arabic language and linguistics*. 5 vols. Leiden – Boston: Brill.

Walther, W.[iebke]. 2010. "Dialect in literature, medieval". In: Meisami & Starkey 2010, 189-190.

Wehr, Hans. 1980. *A dictionary of Modern Written Arabic*. Ed. J. Milton Cowan. Beirut: Librairie du Liban – London: MacDonald & Evans Ltd.

Wehr, Hans. 1985. *Arabisches Wörterbuch für die Schriftsprache der Gegenwart: Arabisch-Deutsch*. Wiesbaden: Harrassowitz Verlag.

Woolard, Kathryn A. & Genovese, Nicholas E. 2007. "Strategic bivalency in Latin and Spanish in early modern Spain". *Language in Society*, 36, 487-509.

Wright, William. 1974. *A grammar of the Arabic language, translated from the German of Caspari, and edited with numerous additions and corrections*. 2 vols. Revised by W. Robertson Smith & M.J. de Goeje. Beirut: Librairie du Liban. [1st edn 1859-1862.]

Youssi, Abderrahim. 1992. *Grammaire et lexique de l'arabe marocain moderne.* Casablanca: Wallada.

Youssi, Abderrahim. 1995. "The Moroccan triglossia: Facts and implications". *International Journal of the Sociology of Language*, 112, 29-43.

Youssi, Abderrahim. 2013. "Impératifs linguistiques, inerties socioculturelles". *Langage et société* 143, 1, 27-40.

Streszczenie

Dialekt marokański języka arabskiego w piśmie

Studium heterografii wariantywnej jakościowej

Podobnie jak dialekty arabskie w innych krajach arabskojęzycznych dialekt marokański, obejmujący w obecnym rozumieniu *koine* bazującą na dialekcie Casablanki, istnieje w sytuacji dyglosji, tj. komplementarności funkcjonalnej dwóch odmian języka. W odróżnieniu od nauczanego w szkołach, znormalizowanego i cieszącego się prestiżem ogólnoarabskiego języka literackiego dialekt, nabywany w sposób naturalny przez dziecko od opiekunów, pozbawiony skodyfikowanych norm i uważany za nieprestiżowy, używany jest typowo w komunikacji prywatnej, nieoficjalnej, prymarnie ustnej, w domenie kultury ludowej, popularnej i świeckiej. Rodzimi użytkownicy dialektu marokańskiego coraz częściej posługują się nim w piśmie, głównie przy użyciu alfabetu arabskiego (poza tym również łacińskiego). Jego postać pisana dominuje w komunikacji elektronicznej i wkracza w sfery reklamy i mediów, a także literatury pięknej ogłaszanej drukiem, dotąd zastrzeżone dla języka literackiego. Ponieważ w odróżnieniu od tego ostatniego dialekt nie posiada norm pisowni (ortografii), osoby piszące w nim teksty (autorzy, redaktorzy) tworzą i stosują, dość niekonsekwentnie, własne systemy graficzne, ustosunkowując się w ten czy inny sposób do pisowni języka literackiego, czyli języka-dawcy systemu graficznego. W konsekwencji złożony z nich makrosystem graficzny dialektu marokańskiego charakteryzuje się dużym stopniem heterografii wariantywnej, tj. reprezentowania danego wyrazu fonetycznego przez różne wyrazy graficzne (heterogramy wariantywne). Zawarty w pracy opis jednego z jej typów – heterografii wariantywnej jakościowej – bazuje na korpusie obejmującym ok. 1900 stron tekstów literackich wydanych drukiem w latach 1991-2012.

Pewną wariantywność wykazuje już sam system pisma języka literackiego, mimo jego znacznej jednolitości w całym świecie arabskim. Ponadto dla niniejszej analizy istotne jest, że pismo to jest w dużym stopniu kontekstowe, tzn. określenie znaczenia i wymowy wielu wyrazów graficznych nie jest możliwe w oderwaniu od kontekstu. Jego grafię kształtuje szereg zasad ogólnych: fonetyczna, fonologiczna, morfologiczna, historyczna, zasada homonimii, zasada ekonomii, zasada estetyczna i inne.

W dotychczasowych badaniach nad grafią dialektu marokańskiego zidentyfikowano dwie główne kształtujące ją zasady (tendencje), w pracy nazywane zasadą zorientowania na język-dawcę i zasadą autoorientacji, która z kolei obejmuje zasadę fonetyczną i zasadę morfologiczną. Wyodrębnić należy też zasadę kontestacji języka-dawcy, polegającą na dystansowaniu się od grafii języka literackiego. Podobnie jak pismo języka-dawcy grafia dialektu marokańskiego jest w dużym stopniu kontekstowa.

Do analizy grafii dialektu marokańskiego w pracy używa się pojęć z obszaru grafetyki, takich jak *graf*, *homografia*, *heterografia*, *wyraz graficzny ciągły*, i grafemiki, np. *heterografia dystynktywna*, *heterografia wariantywna*, *wyraz graficzny nieciągły*, *wariancja*, *grafem*. Analiza skupia się na 42 grafach prymarnych, tj. podstawowych jednostkach opisu, których zbiór obejmuje 28

liter alfabetu arabskiego oraz ich modyfikacje i znak hamzy. W zależności od charakteru elementu odróżniającego wyrazy graficzne heterografia wariantywna może być (i) linearna (spacja vs. brak spacji), (ii) ilościowa (graf vs. brak grafu) albo (iii) jakościowa (wariancja dwóch różnych grafów w tej samej pozycji). Zależnie od relacji między zbiorami wyrazów graficznych, w których może się manifestować wariancja dwóch grafów, w pracy wyróżnia się wariancję wolną, asymetryczną i ograniczoną. W przypadku grafów użytych w dwóch heterogramach wariantywnych na tej samej pozycji, lecz oznaczających dwa różne i niepowiązane zjawiska fonetyczne mamy do czynienia z pseudowariancją.

Jedną z trudności związanych z podjętą w pracy próbą ustalenia inwentarza grafemów dialektu marokańskiego jest zjawisko neutralizacji relacji funkcjonalnej wiążącej niektóre jego grafy. Proponuje się wobec tego kryterium, zgodnie z którym grafem rozumiany jest jako zbiór grafów pozostających ze sobą w określonych typach wariancji, niezdolnych do odróżniania znaczeniowego i fonetycznego wyrazów graficznych.

Do problemów związanych z opisem grafii dialektu marokańskiego należy zjawisko nieokreśloności funkcjonalnej niektórych grafów, polegającej na niemożliwości jednoznacznego przyporządkowania im funkcji fonetycznej lub funkcji innego rodzaju. Inna trudność dotyczy rozróżnienia między nietypowymi formami graficznymi wynikłymi z niestaranności edytorskiej bądź względów typograficznych a nietypowymi formami reprezentującymi tendencję graficzną.

Precyzując zidentyfikowane w dotychczasowych badaniach zasady (tendencje) kształtujące badaną grafię, dla zasady zorientowania na język-dawcę wyróżnia się dwie podzasady: zasadę zorientowania na *formę* języka-dawcy i zasadę zorientowania na *zasadę* języka-dawcy. Zasadę fonetyczną dzieli się na zasadę pisowni fonetycznej według reguł przejętych z języka-dawcy (jego odmiany lokalnej albo powszechnej) i zasadę pisowni fonetycznej według reguł adaptowanych, przy czym adaptacja może być endogenna (tj. z systemu arabszczyzny literackiej – motywowana fonetycznie bądź historycznie) albo egzogenna (tj. spoza niego). Niektóre formy zapisu oparte są na (fałszywej) analogii.

Przedstawiając reprezentatywny wybór blisko 240 par heterogramów wariantywnych jakościowych ilustrujących wariancje między grafami wraz ze wskazaniem zasad odpowiadających za poszczególne formy graficzne, w pracy identyfikuje się elementy grafii dialektu marokańskiego podlegające wariancji jakościowej i wyjaśnia przyczyny tego zjawiska. Z zestawienia wariancji wynika, iż tylko jedną parę grafów, گ ‹ǧ› i ڴ ‹ĝ›, wiąże wolna wariancja poświadczona w korpusie w sposób niemarginalny. 10 par reprezentuje wariancję asymetryczną (z tego 3 pary tylko w niektórych pozycjach w wyrazie graficznym). Tylko 5 grafów nie ulega żadnej wariancji. Wariancja ograniczona, najbardziej nieregularna, okazuje się najbardziej rozpowszechniona, dotyczy bowiem 32 grafów, tj. 75% ich ogółu.

Analiza pozwala również sformułować warunki inwariancji jakościowej dla wyrazów graficznych oraz scharakteryzować wieloznaczność fonetyczną opisywanego systemu pisma. Ponadto na podstawie zaproponowanych kryteriów określono inwentarz 34 grafemów dialektu marokańskiego wraz z ich allografami.

Wśród sygnalizowanych w pracy nowych ścieżek badawczych wartych podjęcia znajdują się m.in. włączenie do opisu dwóch pozostałych typów heterografii wariantywnej: linearnej i ilościowej, poszerzenie korpusu o teksty należące do innych gatunków (prasowe, reklamowe, dydaktyczne) i porównanie pod kątem grafetyki i grafemiki dialektu marokańskiego z językiem literackim oraz innymi dialektami posiadającymi rozwijające się piśmiennictwo, przede wszystkim egipskim, z zastosowaniem zaproponowanego aparatu teoretycznego.

WYDAWNICTWO
NAUKOWE

WYDAWNICTWO NAUKOWE

UNIWERSYTETU IM. ADAMA MICKIEWICZA

W POZNANIU

ul. A. Fredry 10, 61-701 Poznań
tel. 61 829 46 46
e-mail: wydnauk@amu.edu.pl
www.press.amu.edu.pl